Evolutionary Computation for Intelligent Data Analytics
Key Applications in Engineering

Amir H Gandomi

Professor of Data Science at University of Technology Sydney
Distinguished Professor at Óbuda University, Budapest

Robert Bosch Engineering company, Budapest, Hungary
20 November, 2023
AI in Action – An overview

What is Artificial Intelligence (AI) ?

AI: The ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning

A convergence of algorithmic advances, data proliferation, and tremendous increases in computing power and storage has propelled AI from hype to reality

https://www.mckinsey.com/business-functions/mckinsey-analytics/
Leaders in the adoption of artificial intelligence also intend to **invest more in the near future** compared with laggards.

**Future artificial intelligence (AI)-demand trajectory, % change in AI spending over next 3 years**

1 Estimated average, weighted by company size; demand trajectory based on midpoint of range selected by survey respondent.

2 Adopting 1 or more AI technologies at scale or in business core; weighted by company size.

*Source:* McKinsey Global Institute AI adoption and use survey; McKinsey Global Institute analysis
How we use AI in Action

AI tools
- ANN, EI, FL

Prediction
- Classification
- Regression
- Clustering
- etc

Optimization
- Cost
- Quality
- Time
- Risk
- Reliability, etc

Automation
- Decision making
- Smart systems
- Robotics

3/9/2020
Outlines

Evolutionary Intelligence in Engineering

- Modelling
- Optimization
- Monitoring
AI/EI for Engineering Modelling
AI-based Data-Driven based Modelling Tools

Discovering patterns

Data Mining → Patterns

Modeling Tools

Artificial Intelligence

Statistics

Neural Network

Fuzzy Logic

Support Vector Machine

Genetic Programming

7/3/2020
AI-based Predictive Data Analytic Tools

Look deep into nature, and then you will understand everything better

Albert Einstein

Nature is the source of all true knowledge.

Leonardo da Vinci

Evolution

Evolution Philosophy
Jalal-Din M. Rumi (Mevlâna)
13th Century
“I died as a mineral and became a plant,
I died as plant and rose to animal,
I died as animal and I was Man.”

Theory of Evolution based on Natural Selection
Charles R. Darwin
1859
“On the Origin of Species”

Evolutionary Search & Learning
Alan M. Turing
1950
“...there is the genetic or evolutionary
search by which a combination of genes is
looked for, the criterion being the survival
value”
What is Genetic Programming

A Software with Evolution as the programmer!

- Many problems can be solved by genetic programming!
- Most popular for predictive data analytics

Reference:
Finding the Model Structure

In Genetic Programming:

- The Structure is found via Evolution
- Pre-defined structure is not required
- Distinguished feature from other machine learning methods
- It can model the behaviour without any prior assumptions

NASA Communication antennas
On the ST-5 mission (2006)

Jason D. Lohn, Gregory S. Hornby and Derek S. Linden,
Simplicity and Explainability


Inherently Making the Selections!

After cleaning data
Feature selection
Model selection
Parallel Processing in Genetic Programming

Ex. 1.1: Response of Self-Centering Concentrically Braced Frames

Ex. I.1: Formulation of each Record’s Response

\[ \theta = f(\text{Structural Design, Intensity Measures}) \]

### Intensity Measures:
- Elastic spectral acceleration
- Elastic spectral acceleration
- Elastic spectral velocity
- Elastic spectral displacement
- Peak ground acceleration
- Peak ground velocity
- Peak ground displacement
- Cumulative absolute velocity
- Cumulative absolute displacement
- Arias intensity
- Velocity intensity
- Root mean square acceleration
- Characteristic intensity
- Strong ground motion duration

### Structural Design:

<table>
<thead>
<tr>
<th>Geometrical</th>
<th>Mechanical</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b, \text{ ft (m)} )</td>
<td>( h, \text{ ft (m)} )</td>
</tr>
<tr>
<td>22.5 (6.9)</td>
<td>52.5 (16)</td>
</tr>
<tr>
<td>30 (9.1)</td>
<td>77.5 (23.6)</td>
</tr>
<tr>
<td>40 (12.2)</td>
<td>102.5 (31.2)</td>
</tr>
</tbody>
</table>
SC-CBF Parameters

Office buildings
Stiff soil site
Los Angeles, CA
Nonlinear Dynamic Analysis

1) 75 SC-CBF System are designed
2) 30 earthquake records in DBE level
3) 140 ground motion records used in the FEMA SAC Steel Project

<table>
<thead>
<tr>
<th>Area</th>
<th>FOE</th>
<th>DBE</th>
<th>MCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles, CA</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Boston, MA</td>
<td>X</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Seattle, WA</td>
<td>X</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
Feature Selection: Evolutionary Coefficient

- Best correlation coefficient (R)!
- R: linear relationship

\[ R_e = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i) \left( f_{j,GP}(x_{ij}) - \bar{f}_{j,GP}(x_{ij}) \right)}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \sum_{i=1}^{n} \left( f_{j,GP}(x_{ij}) - \bar{f}_{j,GP}(x_{ij}) \right)^2}} \]

- \( f_{j,GP} \): Transformed and correlated \( x_j \)

### Feature Selection: Evolutionary Coefficient

<table>
<thead>
<tr>
<th>IM</th>
<th>Symbol</th>
<th>$R^2$</th>
<th>$R_a^2$</th>
<th>$\uparrow$ (%)</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic spectral acceleration</td>
<td>$S_a(T)$</td>
<td>0.5589</td>
<td>0.7975</td>
<td>42.7</td>
<td>3</td>
</tr>
<tr>
<td>Elastic spectral acceleration</td>
<td>$S_a(2T)$</td>
<td>0.6709</td>
<td>0.8680</td>
<td>29.4</td>
<td>2</td>
</tr>
<tr>
<td>Elastic spectral velocity</td>
<td>$S_v$</td>
<td>0.5560</td>
<td>0.7938</td>
<td>42.8</td>
<td>4</td>
</tr>
<tr>
<td>Elastic spectral displacement</td>
<td>$S_d$</td>
<td>0.5147</td>
<td>0.7761</td>
<td>50.8</td>
<td>5</td>
</tr>
<tr>
<td>Peak ground acceleration</td>
<td>PGA</td>
<td>0.4190</td>
<td>0.5359</td>
<td>27.9</td>
<td>10</td>
</tr>
<tr>
<td>Peak ground velocity</td>
<td>PGV</td>
<td>0.7765</td>
<td>0.9022</td>
<td>16.2</td>
<td>1</td>
</tr>
<tr>
<td>Peak ground displacement</td>
<td>PGD</td>
<td>0.5181</td>
<td>0.7222</td>
<td>39.4</td>
<td>6</td>
</tr>
<tr>
<td>Cumulative absolute velocity</td>
<td>CAV</td>
<td>0.1890</td>
<td>0.5694</td>
<td>201.3</td>
<td>11</td>
</tr>
<tr>
<td>Cumulative absolute displacement</td>
<td>CAD</td>
<td>0.4036</td>
<td>0.6729</td>
<td>66.7</td>
<td>7</td>
</tr>
<tr>
<td>Arias intensity</td>
<td>$I_A$</td>
<td>0.1461</td>
<td>0.6612</td>
<td>352.6</td>
<td>8</td>
</tr>
<tr>
<td>Velocity intensity</td>
<td>$I_v$</td>
<td>0.4233</td>
<td>0.6454</td>
<td>52.5</td>
<td>9</td>
</tr>
<tr>
<td>Root mean square acceleration</td>
<td>$A_{rms}$</td>
<td>0.2858</td>
<td>0.3235</td>
<td>13.2</td>
<td>13</td>
</tr>
<tr>
<td>Characteristic intensity</td>
<td>$I_c$</td>
<td>0.2053</td>
<td>0.3305</td>
<td>61.0</td>
<td>12</td>
</tr>
<tr>
<td>Strong ground motion duration</td>
<td>$T_D$</td>
<td>0.0216</td>
<td>0.0881</td>
<td>307.9</td>
<td>14</td>
</tr>
</tbody>
</table>
Formulation of each Record’s Response

**Multi-Objective Strategy**

\[ \ln(\theta) = 25.9PGV + 0.615 \ln(\tanh(2S_a(T))) \left( S_a(2T) + \left(\frac{h}{b}\right)^2 \sqrt{F_y}\right) - 1.08 \]

Team Solution: Steel Semi-Rigid Joints

Prediction
- flexural resistance
- initial rotation stiffness

Joint types
- Extended endplate joint
- Welded joint
- Bolted angle joint

### Teams' GP Solution

#### Performance statistics of models for flexural resistance prediction for all element test data

<table>
<thead>
<tr>
<th>Type of Joint</th>
<th>LGP (single solution)</th>
<th>LGP (team solution)</th>
<th>EC3</th>
<th>GP/SA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R$</td>
<td>$MAE$</td>
<td>$MSE$</td>
<td>$R$</td>
</tr>
<tr>
<td><strong>Bolted Endplate Joint</strong></td>
<td>0.9986</td>
<td>8.5496</td>
<td>123.68</td>
<td>0.9975</td>
</tr>
<tr>
<td><strong>Welded Joint</strong></td>
<td>0.9819</td>
<td>22.06</td>
<td>1073.5</td>
<td>0.9880</td>
</tr>
<tr>
<td><strong>Bolted Joints with Angles</strong></td>
<td>0.9918</td>
<td>3.2328</td>
<td>25.82</td>
<td>0.9964</td>
</tr>
</tbody>
</table>

#### Performance statistics of models for initial rotation stiffness prediction for all element test data

<table>
<thead>
<tr>
<th>Type of Joint</th>
<th>LGP (single solution)</th>
<th>LGP (team solution)</th>
<th>EC3</th>
<th>GP/SA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R$</td>
<td>$MAE$</td>
<td>$MSE$</td>
<td>$R$</td>
</tr>
<tr>
<td><strong>Bolted Endplate Joint</strong></td>
<td>0.9969</td>
<td>3.2788</td>
<td>59.11</td>
<td>0.9985</td>
</tr>
<tr>
<td><strong>Welded Joint</strong></td>
<td>0.9734</td>
<td>9.9467</td>
<td>201.14</td>
<td>0.9735</td>
</tr>
<tr>
<td><strong>Bolted Joints with Angles</strong></td>
<td>0.9901</td>
<td>1.88</td>
<td>5.5</td>
<td>0.9923</td>
</tr>
</tbody>
</table>

Multiple Regression in GP

Multi-Gene Symbolic regression (Searson et al. 2010)

\[ Y = d_0 + d_1(A + C)\cos\left(\frac{7}{B}\right) + d_2\sqrt{(5 + A)e^B} \]

Multiple Regression Genetic Programming ( Arnaldo et al. 2014)
Hybrid GP for Concrete Creep Prediction

\[
\hat{J}(t, t_0) = d_0 + d_1 G_1 + d_2 G_2 + d_3 G_3
\]

\[
G_1 = \left(\frac{w}{c}\right) \frac{\ln(t_e + 2.46)}{f'_c}
\]

\[
G_2 = \ln\left[\left(\frac{w}{c}\right) \sqrt{t_0}\right]
\]

\[
G_3 = \left(\frac{f'_c \cdot t_0}{h^2}\right)^2
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Variables.</th>
<th>Accuracy measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(R^2)</td>
</tr>
<tr>
<td>G-C</td>
<td>5</td>
<td>0.83</td>
</tr>
<tr>
<td>Bažant-Baweja B3</td>
<td>10</td>
<td>0.56</td>
</tr>
<tr>
<td>CEB-FIB MC90</td>
<td>5</td>
<td>0.62</td>
</tr>
<tr>
<td>GL2000</td>
<td>6</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Fused GP for Airfoil Self-Noise Prediction

Fusing GP Models by Adaptive Regression by Mixing (ARM) Algorithm

MOOC Performance Modelling using EEG Data

BIG DATA

3Vs

High-volume;
High-variety; and/or
High-velocity

It cannot be handled by the regular information processing techniques: cost-effective and innovative
Why BIG DATA?

How much data?
- >200 million
- 4 million
- 350,000
- 3.8 million
- 500 hours

Moore's law

Genetic Programming for Big Data Analytics (Current)

Divide-and-conquer
- find an efficient and systematic divide-and-conquer strategy, such as information theory
- structure initializations for GP

GP development
- extend GP with a multi-gene structure
- couple GP with other machine learning and/or regression analysis
- extend traditional evolutionary operators and define new operator(s) for the newly developed GP
Multi-stage genetic programming

\[ f(X) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) + f_{int}(X) = \sum_{i=1}^{n} f_i(x_i) + f_{int}(X) \]

\[ f_2(x_2) = f(X) - f_1(x_1) \]

\[ f_3(x_3) = f(X) - f_1(x_1) - f_2(x_2) \]

\[ \vdots \]

\[ f_n(x_n) = f(X) - f_1(x_1) - f_2(x_2) - \cdots - f_{n-1}(x_{n-1}) \]

\[ f_{int}(X) = f(X) - \sum_{i=1}^{n} f_i(x_i) \]

MSGP for Big Data

MSGP for Classification:
Soil Liquefaction modelling

Stage 1:
\[ F_1 = \cos((\arctan(((q_c^2 + 8.372)/(q_c - 8.372)))^2))^3 \]
Stage 2:
\[ F_2 = (-R_f + 1.393)/(R_f + (5.281/R_f)) \]
Stage 3:
\[ F_3 = \sin(-8.297\sigma_v^2 + -6.012 - \sigma_v)/-8.297 \]
Stage 4:
\[ F_4 = ((\arctan(\cos(\sigma_v^3)))^3(\arctan(\cos(\sigma_v))^3)) \]
Stage 5:
\[ F_5 = \arctan(\arctan(((0.0102 - 0.1011/a_{max}) + 0.466))) \]
Stage 6:
\[ F_6 = 0.034 \sin\left(\left(\left(M_w^2\right) - (1.589 - M_w)\right)\right) \]
Stage 7 (interaction):
\[ F_{int} = \cos(((((\cos(M_w)a_{max})(1.534 - M_w + 5.936)) \exp(M_w))^3)/a_{max}))/M_w - 1.534]) \]

Figure: ROC curves

Advancing Genetic Programming via Information Theory

\[
\begin{bmatrix}
G_{11} & G_{21} & \cdots & G_{n1} \\
G_{1m} & G_{2m} & \cdots & G_{nm}
\end{bmatrix}
\times
\begin{bmatrix}
\beta_{11} & \beta_{12} & \cdots & -\beta_{1m} \\
\beta_{n1} & \beta_{n2} & \cdots & -\beta_{nm}
\end{bmatrix}
+ \begin{bmatrix}
\beta_{01} \\
\beta_{0m}
\end{bmatrix}
= \begin{bmatrix}
y_{1pr} \\
y_{mpr}
\end{bmatrix}
\]

<table>
<thead>
<tr>
<th>Population size</th>
<th>Method</th>
<th>WIR</th>
<th>WIW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>MR</td>
<td>0.1012</td>
<td>0.0992</td>
</tr>
<tr>
<td></td>
<td>GP</td>
<td>0.1084</td>
<td>0.1048</td>
</tr>
<tr>
<td></td>
<td>RRGP</td>
<td>0.1012</td>
<td>0.1094</td>
</tr>
<tr>
<td></td>
<td>MRGP-TC</td>
<td>0.1172</td>
<td>0.0958</td>
</tr>
<tr>
<td></td>
<td>MRGP-SumT</td>
<td>0.1115</td>
<td>0.0959</td>
</tr>
<tr>
<td></td>
<td>MRGP-MDL</td>
<td>0.1120</td>
<td>0.0960</td>
</tr>
<tr>
<td></td>
<td>MRGP-SMDL</td>
<td>0.1347</td>
<td>0.0961</td>
</tr>
<tr>
<td>100</td>
<td>MRGP-TC</td>
<td>0.1085</td>
<td>0.0961</td>
</tr>
<tr>
<td></td>
<td>MRGP-SumT</td>
<td>0.1076</td>
<td>0.0963</td>
</tr>
<tr>
<td></td>
<td>MRGP-MDL</td>
<td>0.1030</td>
<td>0.0964</td>
</tr>
<tr>
<td></td>
<td>MRGP-SMDL</td>
<td>0.1080</td>
<td>0.0968</td>
</tr>
<tr>
<td></td>
<td>Proposed GP</td>
<td><strong>0.0617</strong></td>
<td><strong>0.0664</strong></td>
</tr>
</tbody>
</table>

AI/EI for Engineering Optimization
Optimization Algorithms

- Mathematical
- Nature-Inspired
  - Evolutionary Computation
  - Swarm Intelligence

- Global Optimization
  - Gradient-free
  - • Complex
  - • Constraints
  - • Other strategies
  - • Random variable(s)
EC in Real-World Problems

Boeing Turbine geometry of 777 GE engine Design:


Merck Pharmaceutical discovered first clinically-approved antiviral drug for HIV:


& so many other companies.
Traditional Algorithms:
- Genetic Algorithm (GA)
- Simulated Annealing (SA)
- Particle Swarm Optimization (PSO)

Recent Algorithms:
- Aquila Optimizer (AO)
- Krill Herd Algorithm (KH)
- Fire Hawk Optimizer (FHO)
- Salp Swarm Algorithm (SSA)
- Interior Search Algorithm (ISA)
- Prairie Dog Optimization (PDO)
- Marine Predators Algorithm (MPA)
- Partial Reinforcement Optimizer (PRO)

Elsevier – IFAC Best Theory Paper Award in EAAI from 2020-2022
Complex Systems: 35-Storey Space Tower

1262 members and 936 degrees of freedom


11/26/2023
Many Objective Evolutionary Optimization

- 18 evolutionary many-objective algorithms are compared against well-known combinatorial problems!
  - knapsack problem,
  - traveling salesman problem,
  - quadratic assignment problem

- 3, 5, and 10 objectives problems are tested

EMO for Water and Environmental Problems

Optimisation of Local Integrated Water Management System. Objectives:

- mean daily piped potable water,
- mean daily site run-off,
- mean daily nominal costs.

Several studies on multi-objective optimization of wave energy converters (locations, layouts, shapes, etc).
EI for Combating COVID-19

**XPRIZE-Cognizant Pandemic Challenge**

F1: daily new cases
F2: stringency of planned interventions

**UTS team:**
I led team **Kangaroos** in this competition

Our team used
- ML to build models
- EMO to optimize the objectives

**Results:**
- We were the only team from Australia to reach the final
- We end up as a top-ten team
- We became one of the Honourable Mention Winners

Customization in Optimization

AI-based Algorithms

• PROs
  – Derivative-free
  – Global
  -- Flexible

  o Heuristic
    o More efficient
      o Convergence
      o Speed

• CONs
  – Slow
Domain Knowledge

Possible Knowledge

- Expert Knowledge
- Information and Mathematical Theories
- Engineering Principles
- Scientific Concepts

Tutorial (most recent):
Semi-Independent Variables (SIV)

\[ x_1 \geq x_2 \geq \ldots \geq x_{i-1} \geq x_i \]
\[ x_2 = x_1 \times p_1 \]
\[ \vdots \]
\[ x_i = x_{i-1} \times p_{i-1} \]
\[ x_{i+1} = x_i \div p_i \]
\[ \vdots \]
\[ x_n = x_{n-1} \div p_{n-1} \]

\( \mathbb{R} = 2^N \)

\( N: \text{number of SIVs} \)

Resource allocation Problem:
• Casting scheduling
• Multi-knapsack problem is NP-hard
• Discrete Variables
• Pop. Size: 60 for all problems

How much is a Billion?
• 4GBytes for a solution, 240GB RAM for a population

AI for Engineering Monitoring and Maintenance
Structural Health Monitoring

- Sungsoo Bridge, Seoul (1994)
- Kobe Earthquake (1995)
- San Francisco-Oakland Bridge (1989)
- Highway Bridge, Minnesota (2007)
Inverse Analysis

Inverse Problem of Engineering Systems

As the system properties are usually unknown we need to do the inverse analysis.

We need to have the Model (F.E. Model)
Coupled Self-Sim & GP

Read Reference Data

FEM-FC
Conducting Newton-Raphson Iterations with Force Boundary Constraint

FEM-DC
Conducting Newton-Raphson Iterations with Displacement Boundary Constraint

Update the Stress Strain Database

Training the ANN model

Self-Sim analysis and derive an ANN based material model

Forward analysis to extract the stress-strain data base

Use GP to explicitly formulate the material model


Gandomi A.H., Yun G.J., "Coupled SelfSim and genetic programming for non-linear material constitutive modelling." Inverse Problems in Science and Engineering, 23(7), 2015
Case Study: 112 bar Space Truss
GP-based Constitutive Model

\[ \sigma = 2 \times 10^{11} \varepsilon \]

\[
\begin{align*}
&f[0]=v[0]; \\
&l0: f[0]^*=0.02; \\
&l1: f[0]=\sqrt{f[0]}; \\
&f[1]-=f[0]; \\
&l2: f[0]-=0.25; \\
&l3: f[0]^*=f[0]; \\
&f[1]+=f[0]; \\
&l4: f[1]+=f[0]; \\
&f[0]+=-f[1]; \\
&l5: f[0]=\text{fabs}(f[0]); \\
&f[0]=-f[1]; \\
&l6: f[1]-=f[0]; \\
&f[0]^*=f[0]; \\
&l7: f[0]^*=f[0]; \\
&f[0]+=-f[1]; \\
&l8: \\
&l9: \\
&\text{return } f[0]; \\
\end{align*}
\]
Forward FE simulation using GP-based constitutive model
EI is only one of the AI tools!

**AI methodologies/Tools:**
- Machine Learning
- Deep Learning
- Natural Language Processing
- Computer Vision
- Predictive Analytics
- Evolutionary Intelligence
- Fuzzy Logic
- Expert Systems
- Robotics
- Signal Processing

3/9/2020
Other studies!

Using other AI based tools such as the following SHM studies:

- Assessment of standing trees and wooden poles using contact–ultrasonic testing and machine learning and convolutional neural network!

- Combining advanced signal processing (VMD and Johansen cointegration) and machine learning approaches for structural health monitoring!
AI Impacts in Engineering

- Enhanced Design and Simulation
- Predictive Maintenance
- Improved Efficiency and Automation
- Quality Control
- Risk Assessment
- Resource Optimization
- Smart Infrastructure
- Environmental Impact
- Supply Chain and Logistics
- Customization
- Safety Enhancements
- Data Management and Decision Making
- Intelligent Monitoring Systems
- Workforce Transformation
- Research and Development
Effect of AI in Performance

25+% increase in speed,
40+% improvement in output quality, and
12+% more tasks completed

Thank you!

Amir H Gandomi

Professor of Data Science at University of Technology Sydney
Distinguished Professor at Óbuda University, Budapest