Comparing various regression methods on ensemble strategies in differential evolution

19th International Conference on Soft Computing, June 26-28, Brno, Czech Republic

Authors: Iztok Fister Jr., Iztok Fister, Janez Brest

July 3, 2013

Agenda

- Motivation
- Introduction
- Background
- Proposed DE algorithm
- Experiments and results
- Conclusion

Motivation

- DE proposes a multitude of strategies
- Which is the best?
- Predicting the best strategy during the run using various regression methods
- Determine the best various regression method

Introduction

- DE(Storn and Price, 1995) is simple, but powerful EA for continuous and combinatorial optimization
- Making decision in computation intelligence and machine learning:
  - Ensemble learning: predicting the regression vector obtained from ensemble strategies (e.g., RF, EXT, GB)
  - Older regression methods: included to complete comparison (e.g., DT, GLM)
- Experiments on suite of five functions taken from literature

Background: DE

- Population based algorithm
- Representation of candidate solution

$$\mathbf{x}_i^{(t)} = (x_{i1}^{(t)}, \ldots, x_{iD}^{(t)}), \quad \text{for } i = 1 \ldots NP, \tag{1}$$

- Mutation

$$\mathbf{u}_i^{(t)} = \mathbf{x}_{r0}^{(t)} + F \cdot (\mathbf{x}_{r1}^{(t)} - \mathbf{x}_{r2}^{(t)}), \quad \text{for } i = 1 \ldots NP, \tag{2}$$

- Crossover

$$w_{i,j} = \begin{cases} u_{i,j}^{(t)} & \text{rand}_j(0,1) \leq CR \vee j = j_{rand}, \\ x_{i,j}^{(t)} & \text{otherwise}, \end{cases} \tag{3}$$

- Selection

$$\mathbf{x}_i^{(t+1)} = \begin{cases} \mathbf{w}_i^{(t)} & \text{if } f(\mathbf{w}_i^{(t)}) \leq f(\mathbf{x}_i^{(t)}), \\ \mathbf{x}_i^{(t)} & \text{otherwise}. \end{cases} \tag{4}$$

Ensemble of DE strategies

Table: Ensemble of DE-strategies

Nr.	Strategy	Expression
1	Best/1/Exp	$x_{i,j}^{(t+1)} = best_j^{(t)} + F \cdot (x_{r1,j}^{(t)} - x_{r2,j}^{(t)})$
2	Rand/1/Exp	$x_{i,j}^{(t+1)} = x_{r1,j}^{(t)} + F \cdot (x_{r2,j}^{(t)} - x_{r3,j}^{(t)})$
3	RandToBest/1/Exp	$x_{i,j}^{(t+1)} = x_{i,j}^{(t)} + F \cdot (best_i^{(t)} - x_{i,j}^{(t)}) + F \cdot (x_{r1,j}^{(t)} - x_{r2,j}^{(t)})$
4	Best/2/Exp	$x_{i,j}^{(t+1)} = best_i^{(t)} + F \cdot (x_{r1,i}^{(t)} + x_{r2,i}^{(t)} - x_{r3,i}^{(t)} - x_{r4,i}^{(t)})$
5	Rand/2/Exp	$x_{i,j}^{(t+1)} = x_{r1,i}^{(t)} + F \cdot (x_{r2,i}^{(t)} + x_{r3,i}^{(t)} - x_{r4,i}^{(t)} - x_{r5,i}^{(t)})$
6	Best/1/Bin	$x_{j,i}^{(t+1)} = best_i^{(t)} + F \cdot (x_{r1,i}^{(t)} - x_{r2,i}^{(t)})$
7	Rand/1/Bin	$x_{j,i}^{(t+1)} = x_{r1,j}^{(t)} + F \cdot (x_{r2,j}^{(t)} - x_{r3,j}^{(t)})$
8	RandToBest/1/Bin	$x_{j,i}^{(t+1)} = x_{i,j}^{(t)} + F \cdot (best_i^{(t)} - x_{i,j}^{(t)}) + F \cdot (x_{r1,j}^{(t)} - x_{r2,j}^{(t)})$
9	Best/2/Bin	$x_{j,i}^{(t+1)} = best_i^{(t)} + F \cdot (x_{r1,i}^{(t)} + x_{r2,i}^{(t)} - x_{r3,i}^{(t)} - x_{r4,i}^{(t)})$
10	Rand/2/Bin	$x_{j,i}^{(t+1)} = x_{r1,i}^{(t)} + F \cdot (x_{r2,i}^{(t)} + x_{r3,i}^{(t)} - x_{r4,i}^{(t)} - x_{r5,i}^{(t)})$

Background: Ensemble learning methods

- Random forest (RF):
  - Constructing many decision trees during the training
  - Classification and regression
- Extremely randomized trees (EXT):
  - Similar to RF
  - The same set is used to train all trees
- Gradient boosting (GB):
  - Natural handling of mixed data type
  - Predictive power
  - Robustness to outliers

Background: Other regression methods

- Decision tree (DT):
  - Creating a model (training)
  - Predicting the value of target variable
  - Learning simple decision rules inferred from the data features
- Generalized linear model
  - Ridge regression (Tikhonov regularization)
  - Solving a regression model, where the loss function is the linear least squared function

The proposed DE algorithm 1/2

Algorithm 1 The proposed DE algorithm

```
Initialize the DE population x_i = (x_i1, ..., x_iD) for i = 1 ... NP
repeat
  for i = 1 to i ≤ NP
    Create test set T on vector x_i using ensemble strategies ES;
    Create validation set v_i by applying strategy 'rand/1/bin' on vector x_i;
    Build regression vector r_i by applying the regression method using T and v_i;
    if (f(r_i) < f(x_i)) Insert r_i into Q;
    else Insert x_i into Q;
    end if
  endfor
  P = Q;
until (Termination condition meet)
```

The proposed DE algorithm 2/2

Algorithm 2 Create test set function

```
T = ∅;
forall s ∈ ES
    Create vector t using ensemble strategy s on vector x_i;
    Add vector t to test set T;
endfor
```

Experiments

- Goal: how various regression methods influence the performance of original DE algorithm
- Algorithms in test: DE, DE+RF, DE+EXT, DE+GB, DE+DT, DE+LM
- DE parameters:
  - $F = 0.5$, $CR = 0.9$
  - $NP = 10$
  - $FEs = 10{,}000$
  - 25 runs
- RF and EXT regression methods used 40 estimators

Test suite

Table: Test suite

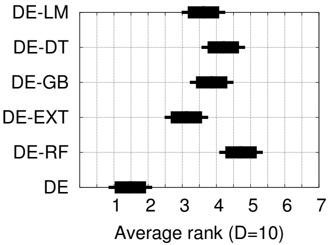
f	Function	Definition	Range
f_1	Rosenbrock	$F(\mathbf{x}) = \sum_{i=1}^{D-1} 100\,(x_{i+1} - x_i^2)^2 + (x_i - 1)^2$	$-15.00 \leq x_i \leq 15.00$
f_2	Rastrigin	$F(\mathbf{x}) = n * 10 + \sum_{i=1}^{D}(x_i^2 - 10\cos(2\pi x_i))$	$-15.00 \leq x_i \leq 15.00$
f_3	Sphere	$F(\mathbf{x}) = \sum_{i=1}^{D} x_i^2$	$-100.00 \leq x_i \leq 100.00$
f_4	Griewangk	$F(\mathbf{x}) = -\prod_{i=1}^{D} \cos\left(\frac{x_i}{\sqrt{i}}\right) + \sum_{i=1}^{D} \frac{x_i^2}{4000} + 1$	$-600 \leq x_i \leq 600$
f_5	Ackley	$F(\mathbf{x}) = \sum_{i=1}^{D-1} \left(20 + e^{-20} e^{-0.2\sqrt{0.5(x_{i+1}^2 + x_i^2)}} - e^{0.5(\cos(2\pi x_{i+1}) + \cos(2\pi x_i))}\right)$	$-32.00 \leq x_i \leq 32.00$

Results

Alg.	D	Value	f_1	f_2	f_3	f_4	f_5
DE	10	Best	3.30E+003	6.70E+001	1.70E+002	5.30E-004	2.80E+000
		Worst	4.80E+006	3.90E+002	4.00E+003	3.80E-001	1.60E+001
		Mean	1.00E+006	1.90E+002	1.50E+003	5.60E-002	1.10E+001
		Median	4.20E+005	1.70E+002	1.30E+003	2.70E-002	1.10E+001
		StDev	1.30E+006	9.30E+001	1.00E+003	8.30E-002	2.70E+000
DE+RF	10	Best	8.79E+000	0.00E+000	0.00E+000	0.00E+000	4.44E-016
		Worst	8.89E+000	0.00E+000	0.00E+000	0.00E+000	4.44E-016
		Mean	8.87E+000	**0.00E+000**	**0.00E+000**	**0.00E+000**	**4.44E-016**
		Median	8.87E+000	0.00E+000	0.00E+000	0.00E+000	4.44E-016
		StDev	2.00E-002	0.00E+000	0.00E+000	0.00E+000	0.00E+000
DE+EXT	10	Best	7.77E-008	1.09E+001	4.44E-042	2.90E-001	3.99E-015
		Worst	4.16E+000	3.19E+001	1.32E-038	2.90E-001	3.99E-015
		Mean	**2.77E+000**	2.33E+001	2.57E-039	2.00E-001	3.99E-015
		Median	2.72E+000	2.45E+001	1.50E-039	2.10E-001	3.99E-015
		StDev	9.90E-001	4.47E+000	2.88E-039	4.90E-002	0.00E+000
DE+GB	10	Best	4.54E+000	1.43E+001	0.00E+000	9.00E-002	4.44E-016
		Worst	7.73E+000	3.23E+001	0.00E+000	3.60E-001	4.44E-016
		Mean	6.43E+000	2.12E+001	0.00E+000	2.30E-001	4.44E-016
		Median	6.61E+000	2.07E+001	0.00E+000	2.20E-001	4.44E-016
		StDev	8.00E-001	4.35E+000	0.00E+000	6.00E-002	0.00E+000
DE+GB	10	Best	1.29E+000	1.54E+001	0.00E+000	7.00E-002	4.44E-016
		Worst	4.33E+000	2.80E+001	0.00E+000	3.40E-001	4.44E-016
		Mean	2.81E+000	2.30E+001	**0.00E+000**	2.20E-001	**4.44E-016**
		Median	2.90E+000	2.41E+001	0.00E+000	2.20E-001	4.44E-016
		StDev	8.50E-001	3.62E+000	0.00E+000	6.00E-002	0.00E+000
DE+DT	10	Best	0.00E+000	0.00E+000	0.00E+000	1.10E-001	4.44E-016
		Worst	8.90E+000	0.00E+000	0.00E+000	2.90E-001	4.44E-016
		Mean	7.83E+000	**0.00E+000**	**0.00E+000**	2.10E-001	**4.44E-016**
		Median	8.90E+000	0.00E+000	0.00E+000	2.10E-001	4.44E-016
		StDev	2.89E+000	0.00E+000	0.00E+000	4.00E-002	0.00E+000
DE+LM	10	Best	4.54E+000	1.43E+001	0.00E+000	9.00E-002	4.44E-016
		Worst	7.73E+000	3.23E+001	0.00E+000	3.60E-001	4.44E-016
		Mean	6.43E+000	2.12E+001	**0.00E+000**	2.30E-001	**4.44E-016**
		Median	6.61E+000	2.07E+001	0.00E+000	2.20E-001	4.44E-016
		StDev	8.00E-001	4.35E+000	0.00E+000	6.00E-002	0.00E+000

Friedman tests

Figure: Friedman test



- RF significantly improved results of the EXT and LM regression methods

Conclusion

- DE using various regression methods on ensemble strategies
- Each regression method (i.e., RF, EXT, GB, DT, GLM) outperformed the original DE algorithm
- The best regression method was RF
- These promising results showed that using various regression methods on ensemble strategies on DE could be a good direction of the DE future development