

jGE Experiments

Grammatical Evolution - Symbolic Regression Experiments

Experiment 1 – Regular Run (12/06/2006)

1a. Symbolic Regression Tableau for GE (O'Neill-style)

Objective:	Find a function of one independent variable and one dependent variable, in symbolic form that fits a given sample of 20 (x_i, y_i) data points, where the target function is the quadratic polynomial $x^4 + x^3 + x^2 + x$.
Terminal Operands:	x (the independent variable)
Terminal Operators:	The binary operators +, -, /, *, and -
Fitness cases:	The given sample of the pairs (x_i, y_i) of 20 data points in the interval [-1, +1]. The input data points (x_i) are randomly created and their corresponding output points (y_i) are automatically created by the expression $x^4 + x^3 + x^2 + x$.
Raw Fitness:	The sum, of the absolute values of errors taken over the fitness cases (x_i, y_i) . With the above Raw Fitness the best individuals have lower values. For this reason a kind of Adjusted Fitness is used and assigned to each individual. Adjusted Fitness of an individual i is typically defined as following: $Fa(i) = 1 / (1 + Fs(i))$ where Fs the Standardised Fitness of i. In this case the Adjusted Fitness of an individual i is calculated as following: $Fa(i) = 1 / (1 + Fr(i))$ where Fr the Raw Fitness of i. The fitness value varies from 0 to 1 and Invalid individuals will have Raw Fitness Value 0.
Standardised Fitness:	Same as raw fitness.
Wrapper:	Standard productions to generate a Java Class with a main() method which prints the fitness values in the standard output
Parameters:	Population Size (M) = 500, Maximum Generations (G) = 50, Prob. Mutation (Pm) = 0.01, Prob. Crossover (Pc) = 0.9, Prob. Duplication (Pd) = 0.01, Prob. Pruning (Pp) = 0.01, Codon Size = 8, Selection Mechanism: Steady State GA with Generation Gap (G) = 0.9 Initial Population: Min Codons = 20, Max Codons = 30

1b. BNF Grammar

```

<expr> ::= <expr><op><expr>|<var>
<op> ::= + | - | / | *
<var> ::= x
    
```

1c. Symbolic Regression Results

Run	Generations	Phenotype	Raw Fitness
1	50	$x + x * x + x * x * x * x + x * x * x$	0.9999999999999998
2	37	$x * x * x + x * x * x / x * x * x + x * x + x / x * x$	1.0
3	50	$x * x + x$	0.3116190264841827
4	50	$x * x + x * x * x * x * x + x + x * x * x$	0.9999999999999998
5	50	$x * x + x * x * x * x + x * x * x * x + x$	0.9999999999999989
6	50	$x * x + x$	0.20382484819217844
7	17	$x * x * x + x * x * x * x * x + x * x / x * x + x$	1.0
8	50	$x - x + x + x * x * x * x * x + x * x + x * x * x$	0.9999999999999996
9	50	$x + x * x + x * x + x / x * x * x * x * x * x$	0.3810122373225437

10	50	$x * x * x - x + x * x + x * x + x + x$	0.2627586966696644
----	----	---	--------------------

Experiment 2 – Implied Knowledge with Restricted BNF Grammar (12/06/2006)

2a. Symbolic Regression Tableau for GE (O'Neill-style)

Objective:	Find a function of one independent variable and one dependent variable, in symbolic form that fits a given sample of 20 (x_i, y_i) data points, where the target function is the quadratic polynomial $x^4 + x^3 + x^2 + x$.
Terminal Operands:	x (the independent variable)
Terminal Operators:	The binary operators + and *
Fitness cases:	The given sample of the pairs (x_i, y_i) of 20 data points in the interval [-1, +1]. The input data points (x_i) are randomly created and their corresponding output points (y_i) are automatically created by the expression $x^4 + x^3 + x^2 + x$.
Raw Fitness:	The sum, of the absolute values of errors taken over the fitness cases (x_i, y_i) . With the above Raw Fitness the best individuals have lower values. For this reason a kind of Adjusted Fitness is used and assigned to each individual. Adjusted Fitness of an individual i is typically defined as following: $Fa(i) = 1 / (1 + Fs(i))$ where Fs the Standardised Fitness of i. In this case the Adjusted Fitness of an individual i is calculated as following: $Fa(i) = 1 / (1 + Fr(i))$ where Fr the Raw Fitness of i. The fitness value varies from 0 to 1 and Invalid individuals will have Raw Fitness Value 0.
Standardised Fitness:	Same as raw fitness.
Wrapper:	Standard productions to generate a Java Class with a main() method which prints the fitness values in the standard output
Parameters:	Population Size (M) = 500, Maximum Generations (G) = 50, Prob. Mutation (Pm) = 0.01, Prob. Crossover (Pc) = 0.9, Prob. Duplication (Pd) = 0.01, Prob. Pruning (Pp) = 0.01, Codon Size = 8, Selection Mechanism: Steady State GA with Generation Gap (G) = 0.9 Initial Population: Min Codons = 20, Max Codons = 30

2b. BNF Grammar

```

<expr> ::= <expr><op><expr> | <var>
<op> ::= + | *
<var> ::= x

```

2c. Symbolic Regression Results

Run	Generations	Phenotype	Raw Fitness
1	16	$x * x * x + x * x * x * x + x * x + x * x + x$	1.0
2	1	$x * x * x + x * x * x * x * x + x * x + x$	1.0
3	8	$x * x * x + x * x * x * x * x + x * x + x$	1.0
4	50	$x + x * x * x + x * x * x * x * x + x * x$	0.9999999999999998
5	13	$x * x * x + x * x * x * x + x * x + x$	1.0
6	9	$x * x * x + x * x + x * x * x * x + x$	1.0
7	20	$x * x * x + x * x * x * x * x + x * x + x$	1.0
8	50	$x + x * x + x * x * x + x * x * x * x$	0.9999999999999998
9	10	$x * x * x + x * x * x * x + x * x + x$	1.0
10	9	$x * x * x * x + x * x * x + x * x + x$	1.0