NewtoN SR++

Nicolas Lassabe, Paul Gersberg

Nicolas.lassabe@capgemini.com Paul.Gersberg@capgemini.com

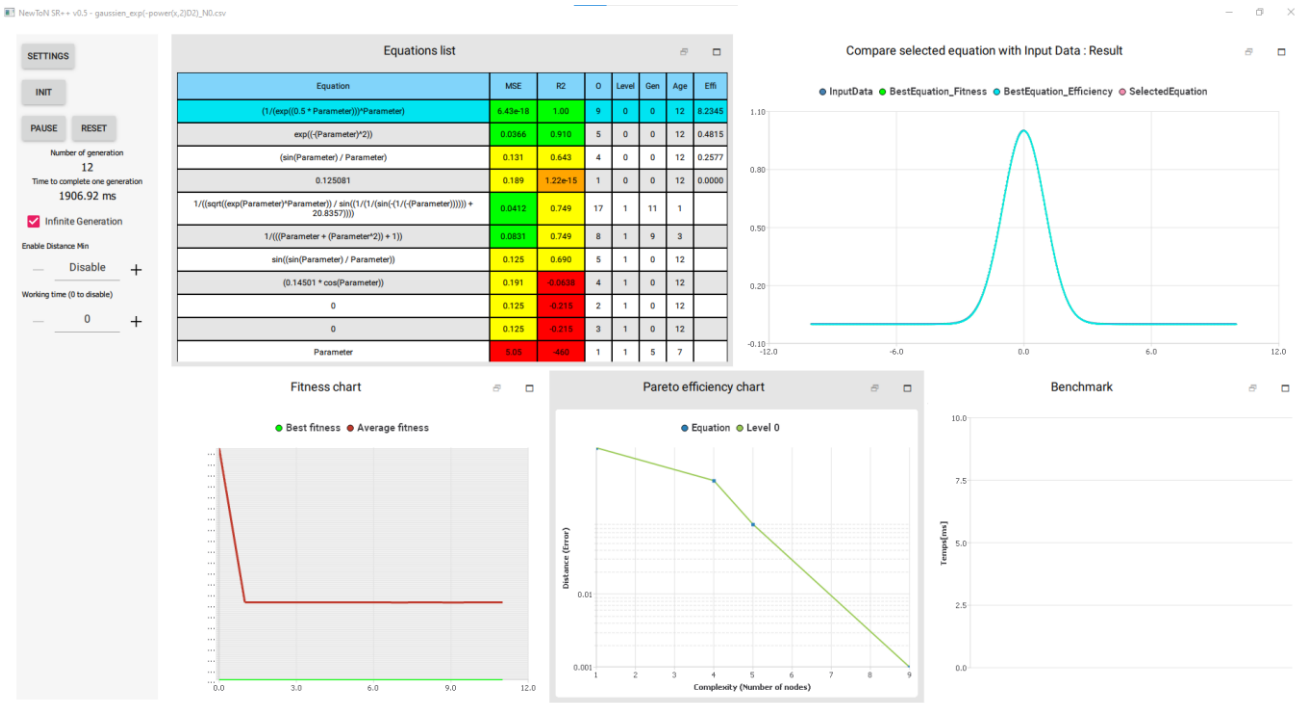
Capgemini Engineering (Toulouse, R&I, 3 chemin de Laporte)

Abstract

Our symbolic regression algorithm uses a multi-branch tree-based genetics algorithm for commutative nodes like addition and multiplication.

The algorithm generates trees which are evaluated according to the method of least squares. They are then classified according to a pareto front taking into account the complexity of the equations and their error. The best equation returned is the one with the best error-to-complexity ratio.
We use 26 types of mutations depending on the type of nodes and leaves.
The constants of the equations are optimized with each generation. We remove all duplicates. Only one equation for the same structure is kept.
Branches that do not contribute to improving the error are deleted. Branches without variant are replaced by constants. The equation are simplified.

For more detail, you can find our C++ source code in the src directory.



[EquationGeneration]
Population=500
MaxDepth=5
MaxNodes=100
NumberPrecision=6
PopulationFitPeriod=3
FitDuration=1ms

[NodeDefaultSelection]
Addition=true
Substraction=true
Multiplication=true
Division=true
Cosinus=true
Sinus=true
Logarithm ln=true
Exponential=true
Power=true
SquareRoot=true
Inverse=true
Arctan=false
Negative=true
Min=false
Max=false
Absolute=false
Tangent=false

[EquationFilter]
FilterFrom=1
SelectedEquations=100
[EquationCrossBreeding]
ParentSelectionStrategy=Tournament
TournamentSize=3
ParentSelectionCriteria=Pareto
CrossBreedingStrategy=Random
CrossBreedingRate=33%
[EquationMutation]
MutationRate=33%
NbMutationPossible=10

[Interface]
MaxParetoSeries=1

[General]
EquationListRefreshRate=1
FitnessRefreshRate=1

ParetoRefreshRate=1
NbOfGenerations=500
EnableDistanceMin=Disable

[EquationDistance]
Distance=R2