

HEAL-Konferenz, 2006-04-06

DI S. Winkler

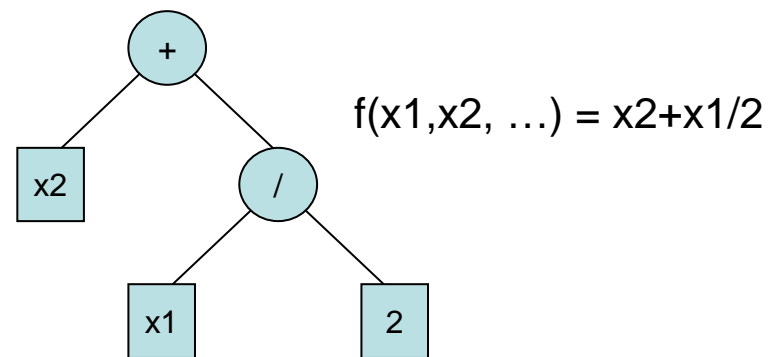
Research Assistant
Austrian Science Fund, Translational-Research-Programm L284
<http://www.fwf.ac.at/>

F&E Competence Center
Upper Austrian University of Applied Sciences
College of Information Technology at Hagenberg, Austria
- and -
Institute for Design and Control of Mechatronical Systems
Johannes Kepler University Linz, Austria

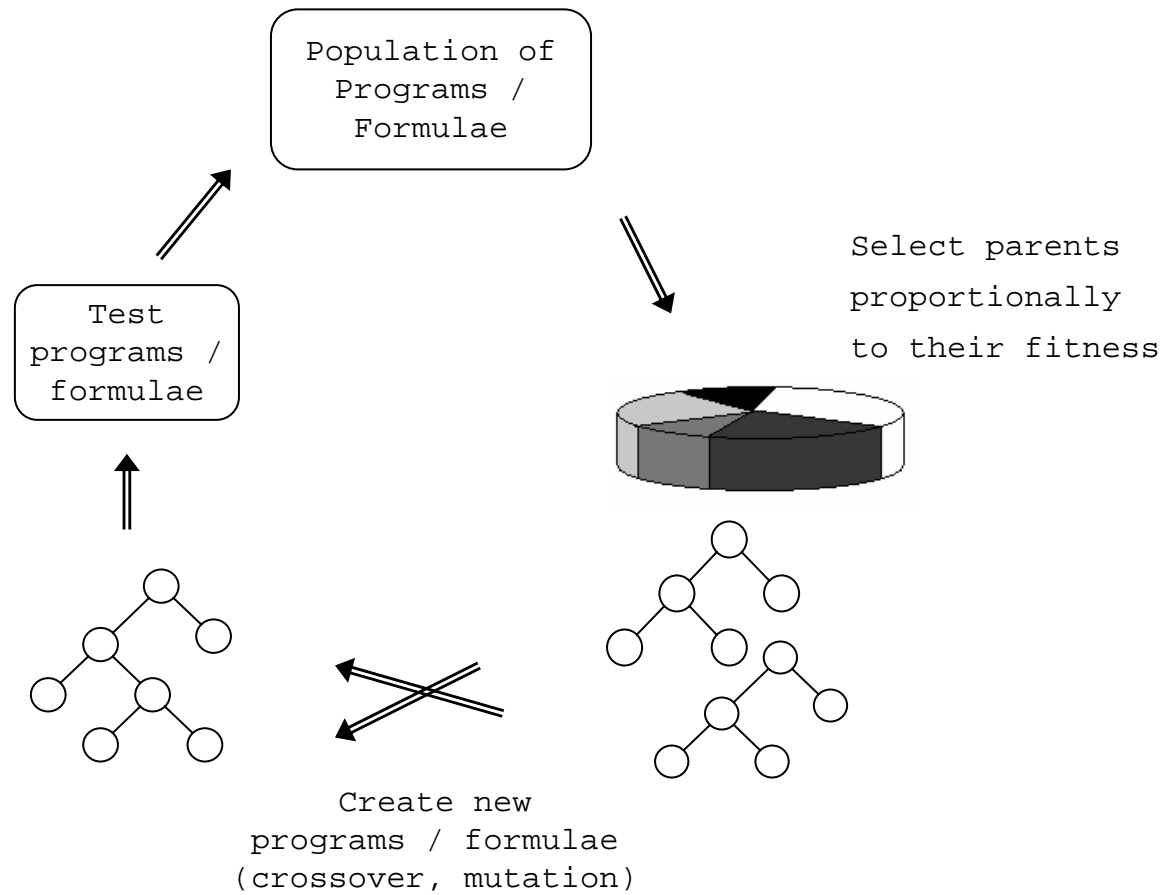
Topics:

- Data-based Structure Identification using GP
- Recent Developments
- Enhanced Algorithmic Aspects: SlimSOP ($1+\lambda$ ES), Tree Pruning
- Recent Publications / Oncoming Conferences
- Demonstration HeuristicModelAnalyzer
- Outlook: GP meets Symbolic Computation

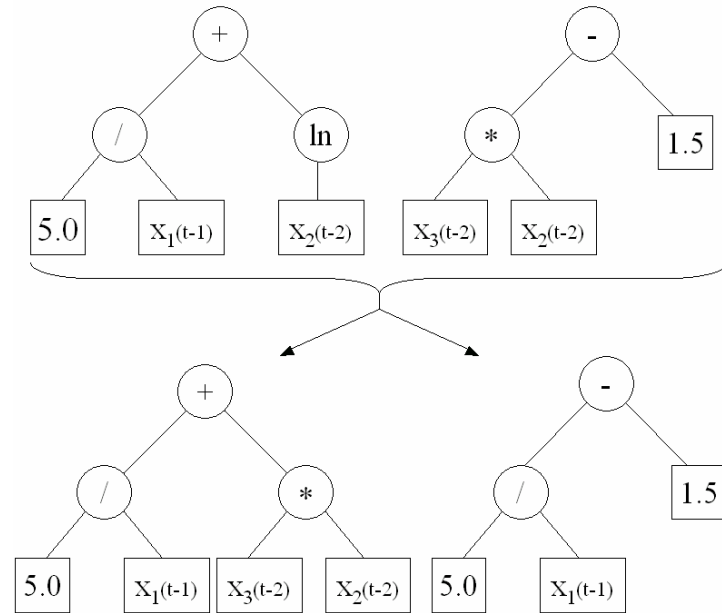
- Goal: Data-Based Generation of Models (Formulae)
- Genetic Programming is used:
 - Automated search for computer programs being able to solve given problems
 - „Automated invention machine“
 - John Koza, 1992: „Genetic Programming: On the Programming of Computers by means of Natural Selection “
 - Based on the theory of Genetic Algorithms
- Formulae are represented as structure trees



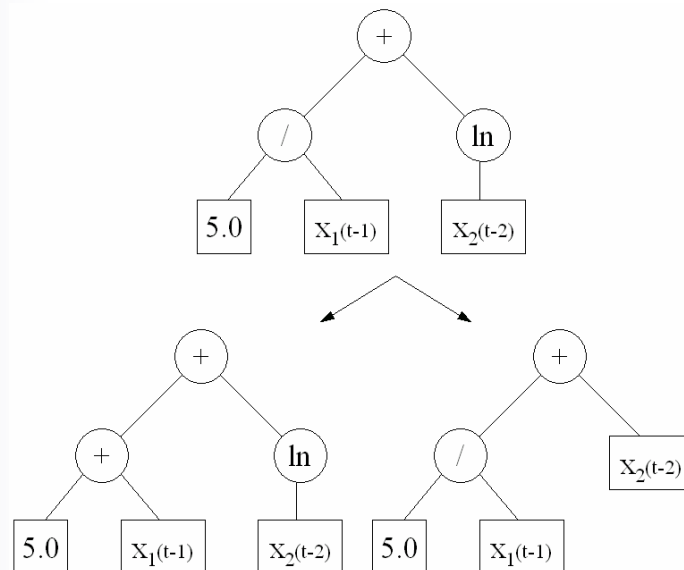
- Data-Based Modeling:
 - Data are given (training data, test data)
 - Arbitrarily many variables and samples
 - One variable represents the target signal
 - E.g.:
 - Diesel engine data: NOx exhaustions or soot as target variables
 - Medical data analysis: Patient classification / diagnosis result as target variable
- Goal: Find formula that describes the target variable(s) as well as possible
- How to measure this?
 - Average squared error between measured and estimated values
 - Classification:
 - Incorporate numbers of correct / incorrect classified samples
 - ROC based
 - Separability, class widths, ...



Crossover:

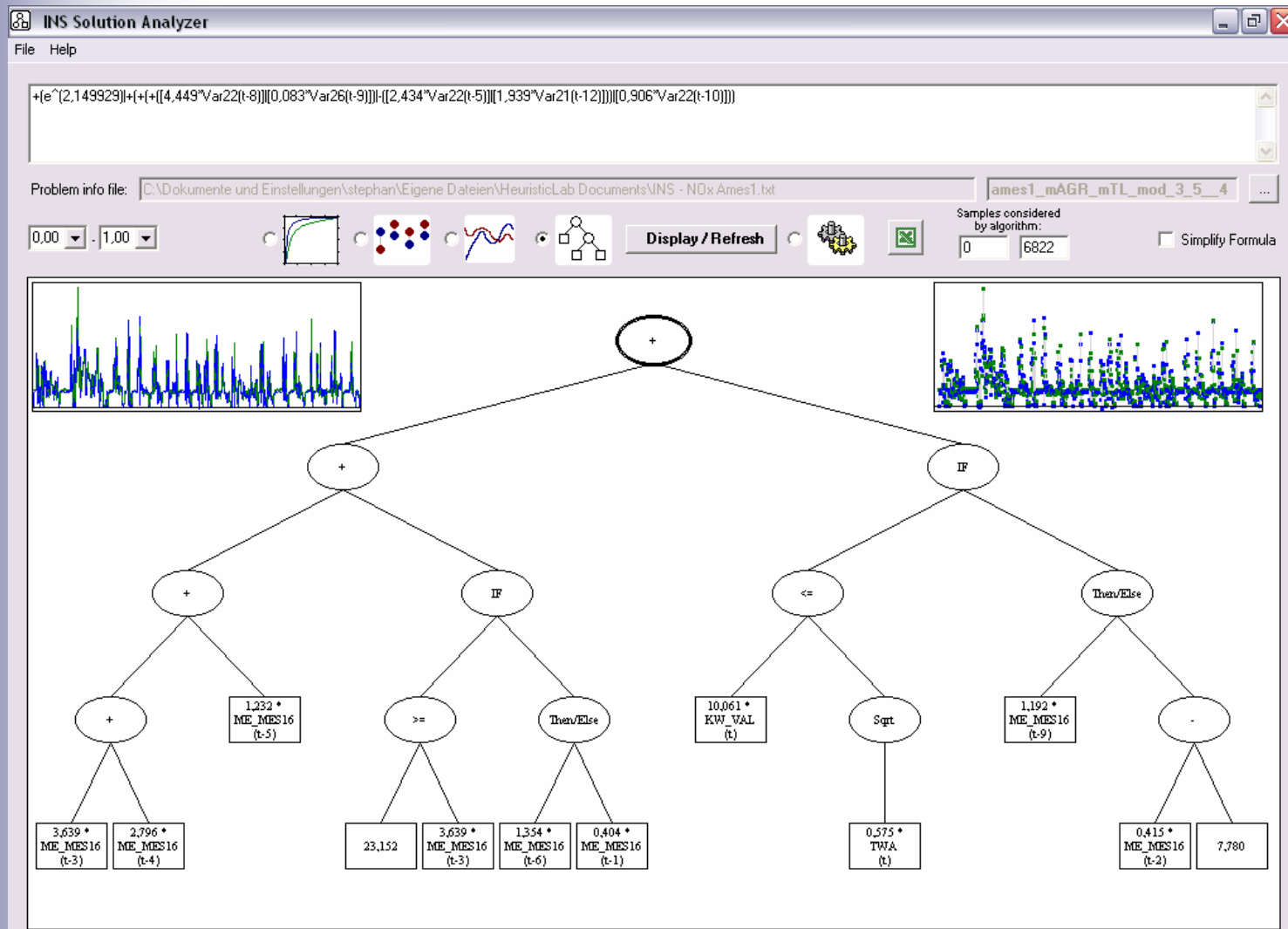


Mutation:



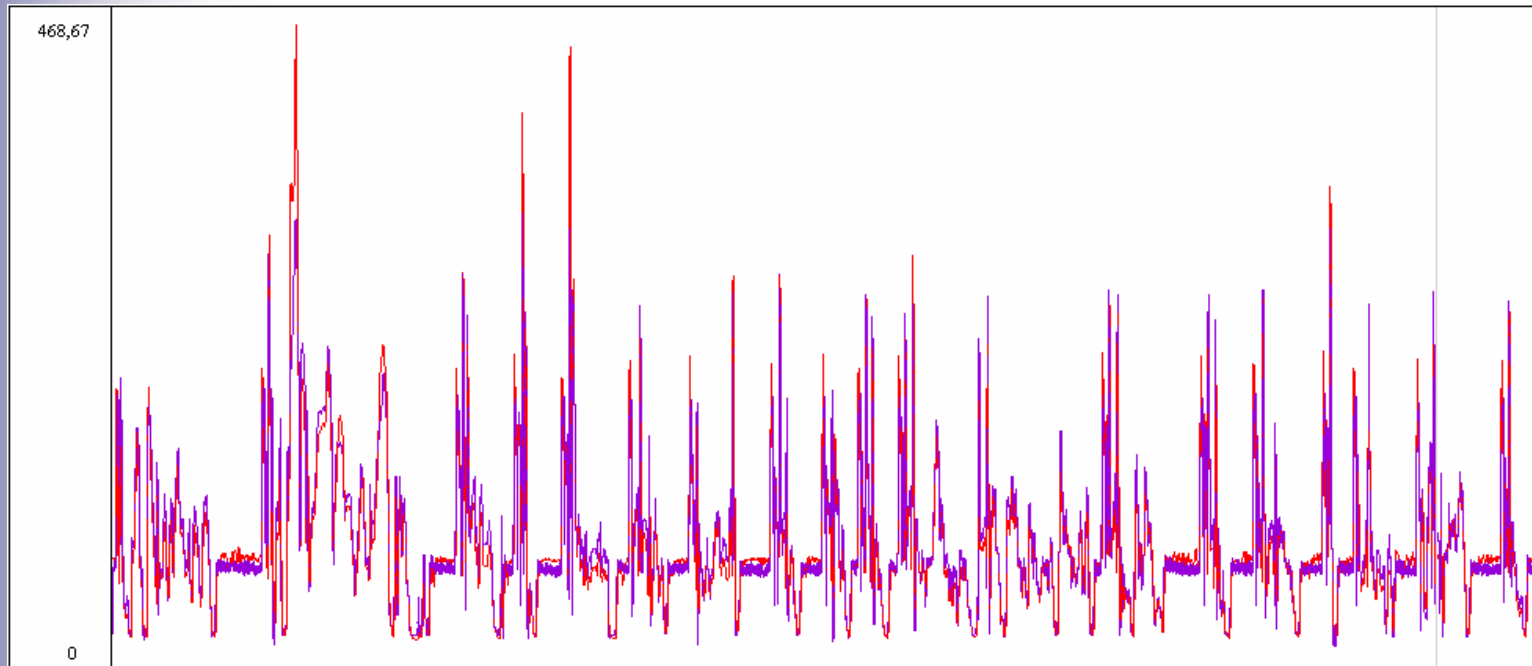
Exemplary Results

DI Stephan Winkler



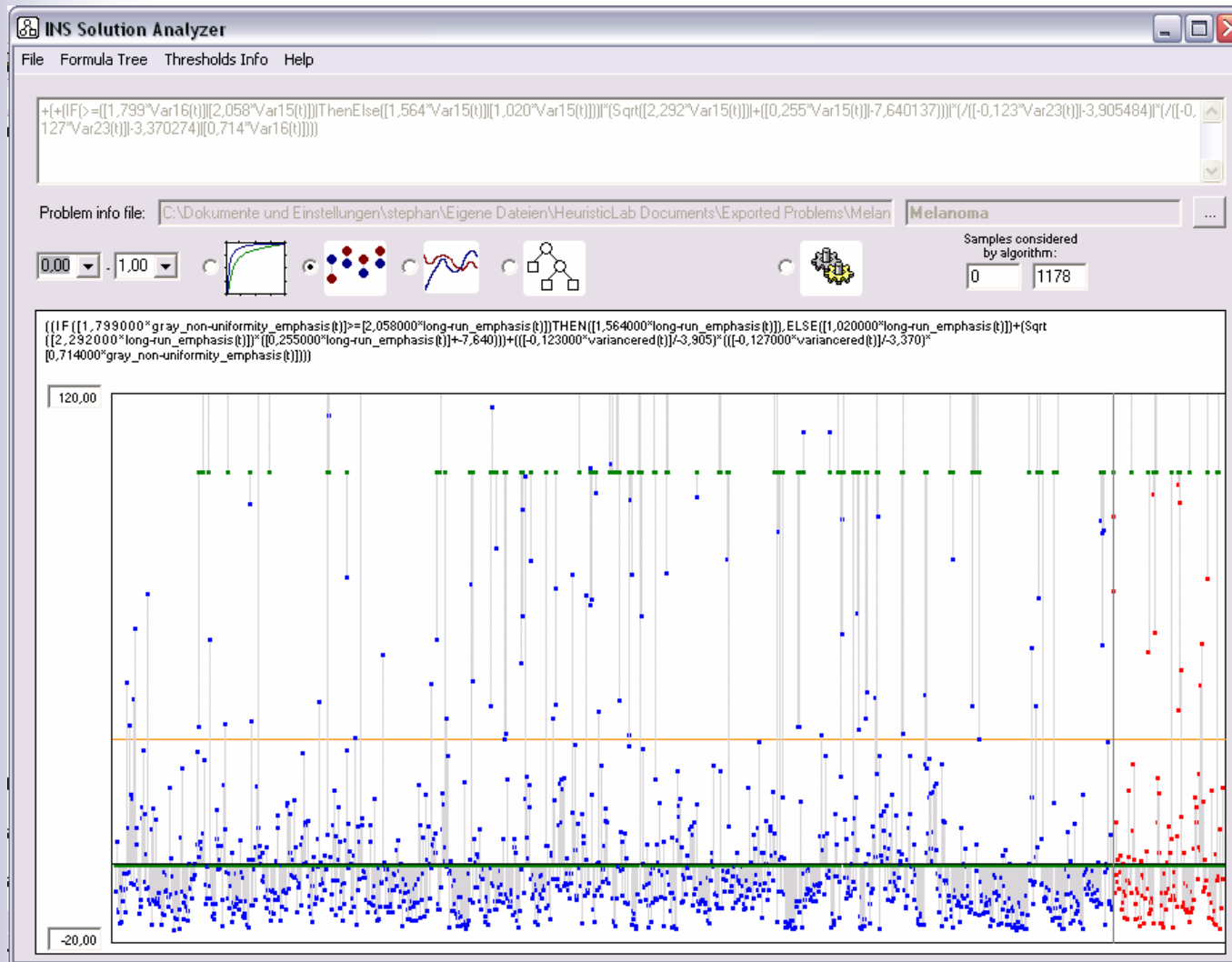
Exemplary Results

DI Stephan Winkler



Exemplary Results

DI Stephan Winkler



- Extensive testing of GP using enhanced algorithmic concepts (SASEGASA) for solving (medical) classification problems
- Extended concepts for analyzing classification models
 - Based on Receiver Operating Characteristics (ROC) Curves
 - For Multi-Class Classification: Sets of MROC curves
- Insertion of additional steps in the GP cycle:
 - ES-based parameter optimization
 - Systematic Tree Pruning
- Development of a tool for analyzing models: The HeuristicModelAnalyzer
 - Includes several methods for analyzing models (standard statistics, MROC graphics, ...)
 - Demonstrates enhanced algorithmic aspects for GP
- Planned: Possibility for user to define „building blocks“, i.e. formula parts that shall be inserted in the initial population as well as during mutation operations.

- **EMCSR (Wien, April):**
 - *18th European Meeting on Cybernetics and Systems Research*
 - *Track: Symposium F: Systems Science in Medicine*
 - [Winkler, Affenzeller, Wagner]:
„Automatic Data Based Patient Classification Using Genetic Programming”
- **NIDISC / IPDPS (Rhodos, April):**
 - *IPDPS: IEEE International Parallel & Distributed Processing Symposium*
 - *NIDISC: Workshop on Nature Inspired Distributed Computing*
 - [Winkler, Affenzeller, Wagner]:
„Advances in Applying Genetic Programming to Machine Learning, Focussing on Classification Problems“
- **GECCO (Seattle, July):**
 - *Genetic and Evolutionary Computation Conference*
 - *Poster acceptance for track Learning Classifier Systems and other Genetics-Based Machine Learning*
 - [Winkler, Affenzeller, Wagner]:
„Sets of Receiver Operating Characteristic Curves and their Use in the Evaluation of Multi-Class Classification“
- **MedGEC (Seattle, July):**
 - *GECCO Workshop on Medical Applications of Genetic and Evolutionary Computation*
 - [Winkler, Affenzeller, Wagner]:
„Using Enhanced Genetic Programming Techniques for Evolving Classifiers in the Context of Medical Diagnosis - An Empirical Study“

→ Demo

- Standard model analysis
- Link to Problem form
- Misclassification weights
- Graphical analysis:
 - Thresholds + info
 - MROC graphics
 - Errors distribution
- Textual analysis
 - Excel export feature
 - Data export
 - Effective classification
 - Thresholds analysis

- Idea: Solve Parametrization Problems using GP
- Example:
 - Given: Curves described by not parametrized functions

Example:

$$f(x, y) = 2x^4 - 3x^2y + y^4 - 2y^3 + y^2$$

- Task: Find (rational) parametrizations
- Example:

$$x(t) = \frac{t^3 - 6t^2 + 9t - 2}{2t^4 - 16t^3 + 40t^2 - 32t + 9},$$

$$y(t) = \frac{t^2 - 4t + 4}{2t^4 - 16t^3 + 40t^2 - 32t + 9}.$$

