Modeling Optimization Algorithm Runtime Behavior and its Applications

Qi Qi, Thomas Weise, Bin Li

qqi@mail.ustc.edu.cn tweise@hfuu.edu.cn binli@ustc.edu.cn

GECCO JULY 2017

Introduction & Approach

Data:

~ sequence of tuples (t_i, q_i), where t_i relating an elapsed amount of t_i of time and to the quality q_i of the best solution discovered with t_i

Models Fitting Quality Φ: $\Phi(M) = 1/n_s(\sum_{i=1:ns}(M(t_i) - q_i)^2/q_i)$

Curve Fitting Method:

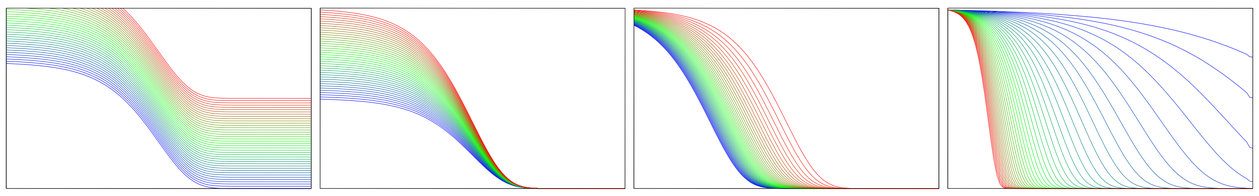
~ Levenberg-Marquardt algorithm
~ intelligent initialization strategy
~ multiple restarts
~ models:

Name	Shortcut	Formula
Decay Model	DCM	$A+B*\exp(C*t^D)$
Logistic Model	LGM	$A+B/(1+\exp(C*\ln(t) + D))$
Exp-Linear Model	ELM	$A+B*\exp(C*\ln(t+D))$
Gompertz Model	GPM	$A+B*\exp(C*\exp(D*t))$

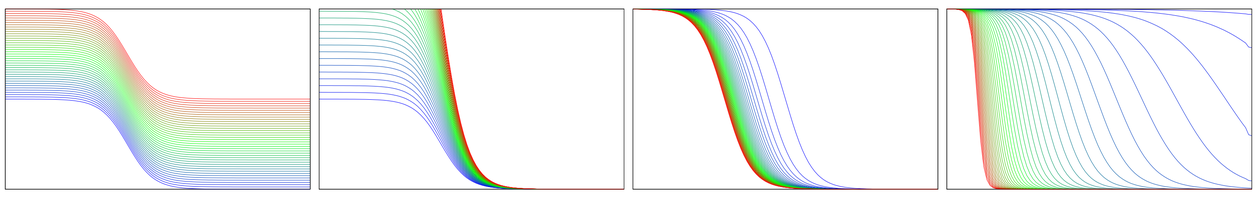
ANNs:

~ represent a much wider range of behavior
~ have significantly more parameters
~ the semantics of these parameters are too complex to be manually interpreted
~ might violate the constraint that any reasonable optimization process will never forget the best solution encountered
~ in our experiment we use feed-forward ANNs with a single hidden layer.
~ implemented in R

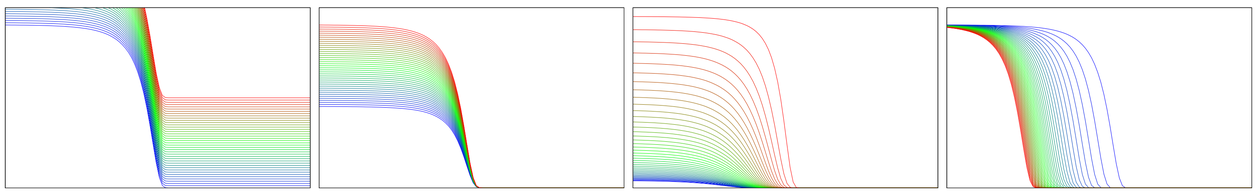
a. Generalized exponential decay (DCMP) $y = M(t) = A + B*\exp(C* t^D)$ for A ϵ [0, 50], B ϵ [50, 100], C ϵ [-0.1, -0.01],D ϵ [0.3, 3]



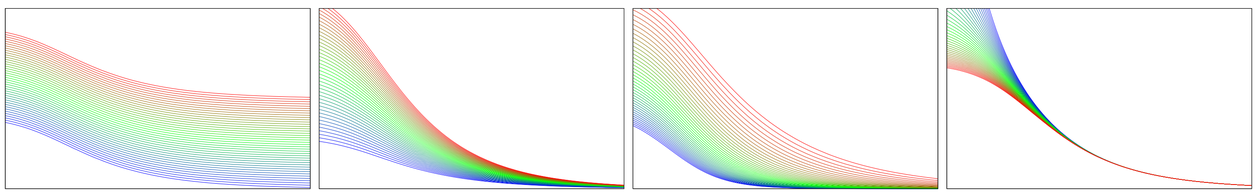
b. Generalized logistc model (LGMP) $y = M(t) = A+B/(1+\exp(C*\ln(t) + D))$ for A ϵ [0, 50], B ϵ [50, 200], C ϵ [0.5, 0.8],D ϵ [0.3, 3]



c. Gompertz function (GPMP) $y = M(t) = A + B*\exp(C*\exp(D*t))$ for A ϵ [0, 50], B ϵ [50, 100], C ϵ [-3.2, -0.05],D ϵ [0.005, 0.15]



d. Exp-linear model (ELMP) $y = M(t) = A + B*\exp(C* \ln(t+D))$ for A ϵ [0, 50], B ϵ [50, 200], C ϵ [-0.1, -0.3],D ϵ [-0.9, 10]



Case Study: MAX-SAT

Problem Introduction & Experiment Setup:

~ six setups with hill climber algorithms: 3 search operator (1, 2, m-bit flip) and whether or not restarts are applied
~ 20 independent runs for each algorithm-instance combination
~ Maximum Satisiability Problem (MAX-SAT) with n variables as case study

Model Parameters vs. Instance Features:

~ analysis relationship between model parameters and algorithm configuration
~ the smaller parameter A the better asymptotic results
~ parameter B has an upward trend for all algorithms with a rising number of variables/clauses of the MAX-SAT instances
~ parameter C and D both have approximately a negative linear association with the instance scale

Algorithm and Instance Classification:

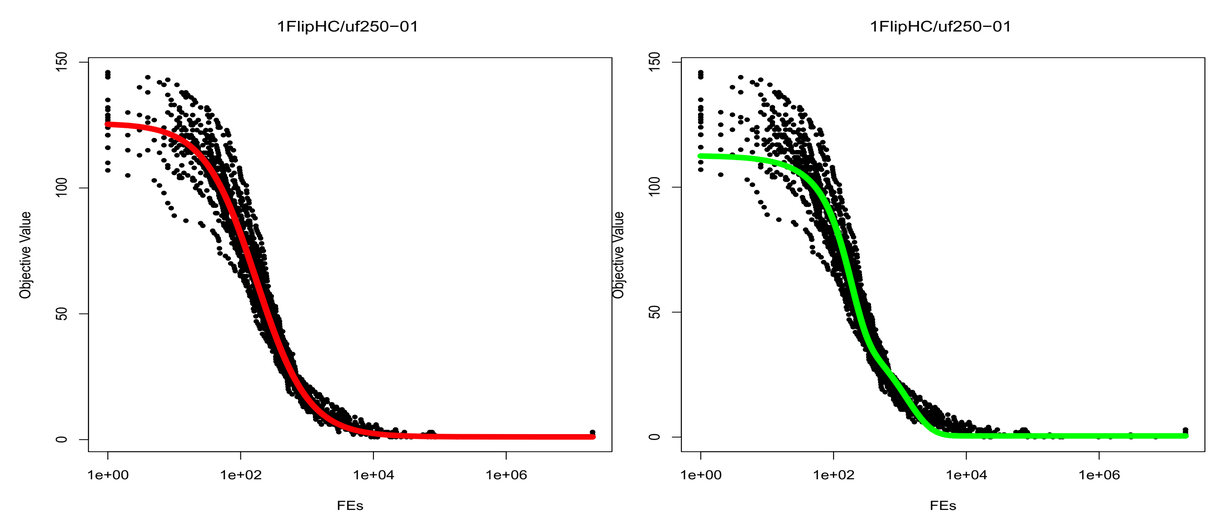
~ samples feature: A, B, C, D and Φ
~ sample class: 6 in total, each algorithm setup as one class.
~ classifier: ANN with backpropagation, Support Vector Machine (SVM) with linear kernel, Gradient Boosting Decision Tree (GBDT)

Model Parameter Prediction:

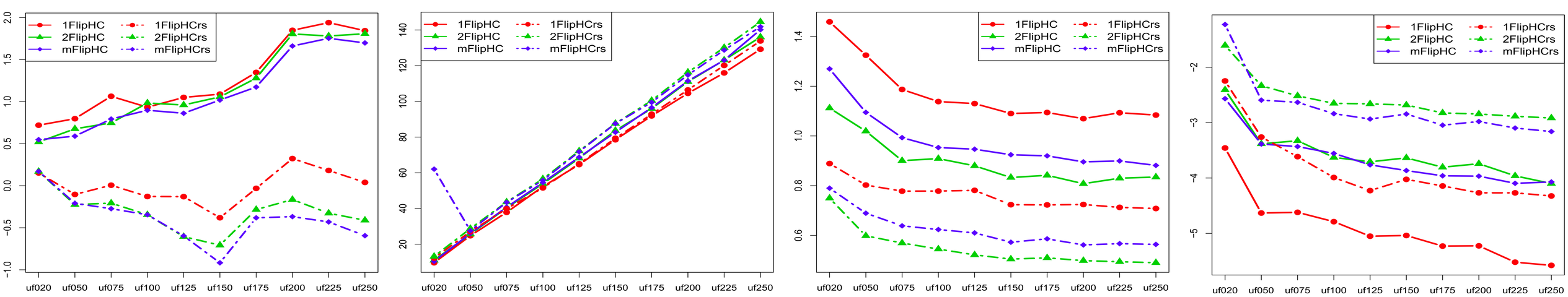
~ the model parameters are clearly related to the instance features.
~ ANN directly predicts model parameters based on instance features.
~ first omit two instance scale parameters results and use ANN predict it and than compared them with the omited true parameters by calculating the Φ
~ ANN wither linear activation function and grid search for weight decay
~ Root Mean Square Error (RMSE) for choosing optimal model

Predicton of Future Progress:

~ representing algorithm behavior as function can compute, for any point in time, which solution quality the algorithm likely has obtained
~ used for prediction future progress of algorithms in the running optimization process.
~ two methods, ANN and Curve Fitting
~ ($train_t$, $test_t$) = (50, 100) stands for predicting the complete algorithm behavior during the first 50 in the future based on the data points collecte



Curve Fittiing result *vs* ANN Fitting result in uf250-01 with 1FlipHC algorithm setup



Trends for the mean values of the four parameters of the fitted LGMP models over he number *n* of variables of the problem instances for the six algorithms.

~ Qi Qi is a graduate student at the University of Science and Technology of China (USTC)
~ Prof. Thomas Weise is the supervisor of Qi Qi and and the director of Institute of Applied Optimization (IAO)
~ Prof. Bin Li is with the School of Information Science of USTC