IRIDIA
Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

Mixed-Effects Modeling of Optimisation Algorithm Performance

Matteo Gagliolo
IRIDIA, ULB, Brussels (BE)
USI, Lugano (CH)

Catherine Legrand
Institut de statistique, UCL,
Louvain-la-Neuve (BE)

Mauro Birattari
IRIDIA, ULB, Brussels (BE)

Abstract

The learning curves of optimisation algorithms, plotting the evolution of the objective vs. runtime spent. can be viewed as a sample of *longitudinal* data. Here we describe *mixed-effects* modeling, a standard technique in longitudinal data analysis, and give an example of its application to algorithm performance modeling.

1 Introduction

Modeling optimisation algorithm performance: how?

- Solve a benchmark of problem instances
- Collect (time,quality) pairs during solution
- Randomized algorithm: repeat with different random seeds

The resulting sample is formed of several sequences of (time,quality) pairs: one sequence for each combination of (algorithm, instance, seed).

How can we model such data?

A statistician would call it longitudinal data: measurements of the same quantity repeated over time, on each of a set of *subjects* (e.g., the weight of an animal, the blood pressure of a patient, . . .).

In our case: each (algorithm, instance, seed) triple is a different subject.

The issue with longitudinal data is within-subject correlation: measurements taken on a same subject cannot be considered independent.

- Single model for all subjects: discards within subject correlation
- Separate model for each subject: discards similarity among subjects
- Solution: mixed effects models

2 Mixed Effects models

Longitudinal data

Sample measurements of a scalar y.

M *subjects*.

n_i *observations* (y_{ij}, t_{ij}) for the i-th subject.

Object of modeling: distribution of y given t.

Mixed effects

- a parametric baseline curve for the set of subjects
- parameters of the curves for each subject: random perturbation of the parameters of the baseline.

A simple linear model

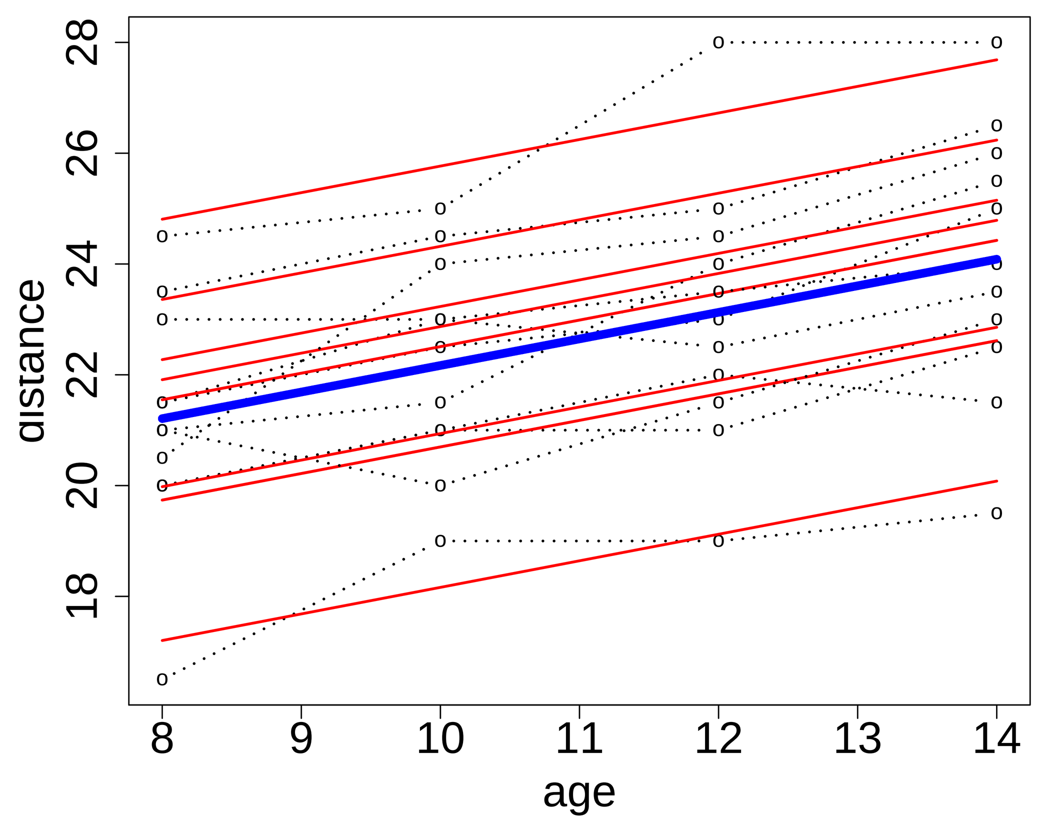
$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i,$$
$$[\mathbf{y}_i]_j = y_{ij}, \ [\mathbf{X}_i]_j = [1, t_{ij}], \ \boldsymbol{\beta} = [\beta_0, \beta_1]^T$$
$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i}),$$

Linear mixed effects (LME)

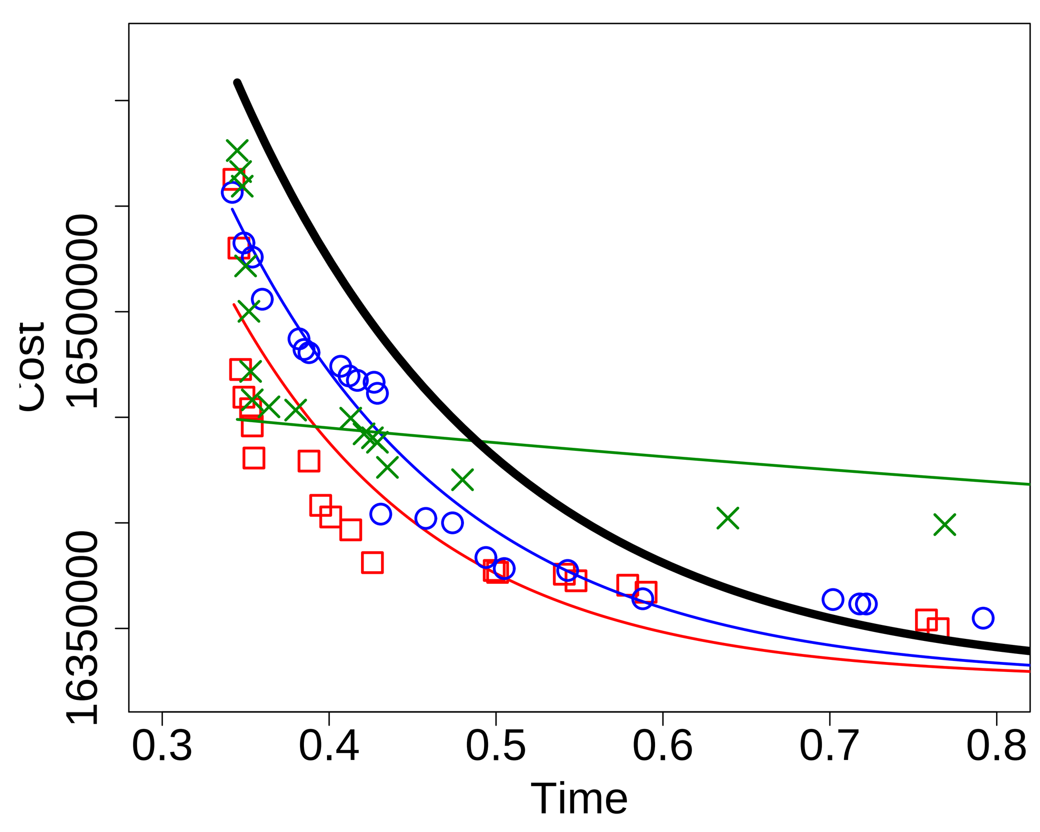
$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,$$
$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i).$$

Nonlinear mixed effects (NLME)

$$y_{ij} = f(\boldsymbol{\phi}_i, t_{ij}) + \epsilon_{ij}$$
$$\boldsymbol{\phi}_i = \boldsymbol{\beta} + \mathbf{B}\mathbf{b}_i$$
$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \ \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Lambda}_i).$$



An example LME model of the Orthodont data.



An example NLME model of ILS-FDD performance on a single instance.

3 Experiments

Scenario

- Algorithm: ILS-FDD (`3-opt`)
- Benchmark: symmetric Euclidean TSP, four groups of 100 instances of size $200, 300, 400, 500$
- NL model: $y = a + be^{-ct}$, $y = (l - l_b)/l_b$

Question: can a be used to estimate the optimum?

Results: statistics of $d = (l_a - l_o)/l_o$, the deviation of the estimate l_a from the actual optimum l_o.

Fig. (a): One NLME model for each instance, based on 25 runs. Each random seed is a subject.

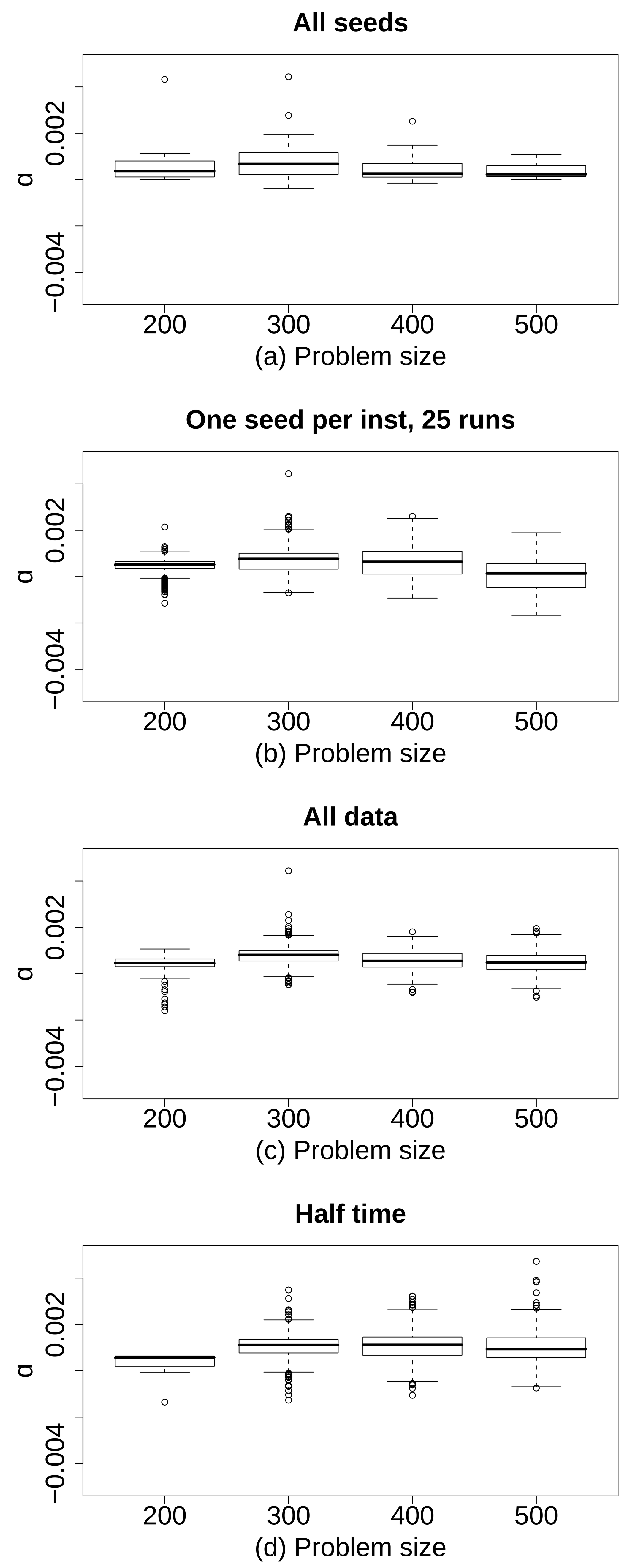
Fig. (b): One model for each group of instances, based on 1 run. Each instance is a subject.

Question: can a be used to predict the optimum?

Idea: simulate a situation in which some instance have been solved, some others are being solved, and the model is used to predict the global optimum on unsolved instances. 50 training and 50 test instances, randomly picked.

Fig. (c): Performance on test instances using all available data (comparison term).

Fig. (d): Performance on test instances discarding observations with t larger than half of the last t recorded.



4 Conclusions

- sound models of algorithm performance
- stationary/time-varying covariates, nested factors
- may be used for smarter stopping criteria
- or algorithm selection/restart strategies

References

G. Fitzmaurice et al. Longitudinal Data Analysis. (2008)
J.C. Pinheiro et al. Mixed Effects Models in S and S-Plus. (2002)

The first author was supported by the SNF (CH).