Efficient Parameter Importance Analysis via Ablation with Surrogates

André Biedenkapp, Marius Lindauer
Katharina Eggensperger, Frank Hutter
University of Freiburg

Chris Fawcett
Holger Hoos
University of British Columbia

Albert-Ludwigs-Universität Freiburg

In a Nutshell

- **Algorithm configuration** often necessary to achieve peak performance (e.g., in MIP, AI Planning, SAT and ASP)
- Costly **parameter importance analysis** to understand which parameter changes are responsible for performance improvements
- Reducing the cost of **ablation analysis** by using **predictions from empirical performance models** instead of real algorithm runs
- **Speed-up factors between 33 and 14 727** in comparison to ablation analysis with racing

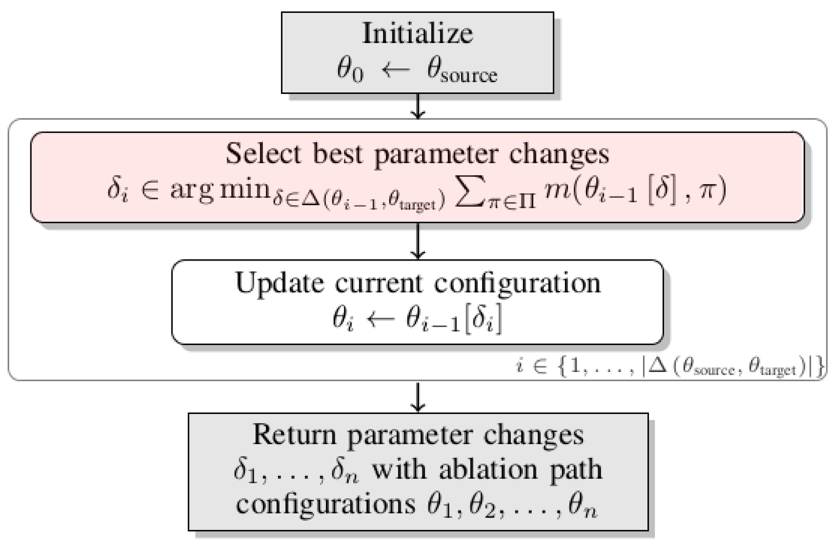
Related Work

- **Ablation** [Fawcett and Hoos. MIC 2013, Journal of Heuristics 2016]: introduced ablation analysis + racing-based extension to reduce time
- **fANOVA** [Hutter et al. ICML 2014]: parameter importance with functional ANOVA using random forests
- **Empirical performance models** [Hutter et al. AIJ 2014]: predict performance of parameter configuration on given instance
- **Using surrogates for efficient hyperparameter optimization benchmarks** [Eggensperger et al. AAAI 2015]: using predictions from empirical performance models instead of real algorithm runs

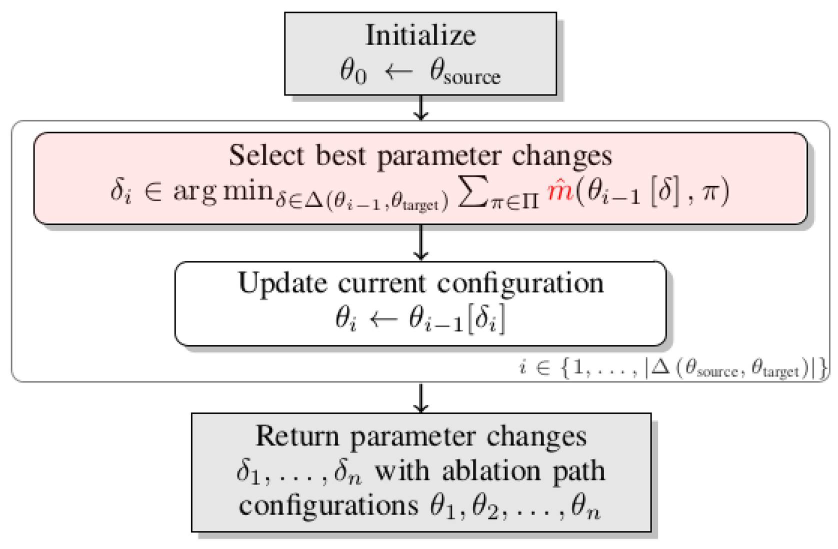
Notation

- θ_{source} e.g., default configuration
- θ_{target} e.g., optimized configuration
- m performance metric (e.g., running time)
- Π instance set (e.g., SAT or MIP instances)
- $\Delta(\cdot,\cdot)$ parameter values differing between two configurations
- κ running time cutoff
- $\hat{m}: \Theta \times \Pi \to \mathbb{R}$ empirical performance model

Ablation Analysis



Efficient Analysis



Our Approach

1. **Gather training data**
   - During configuration, lot of data is generated → Focus on high performance regions
2. **Train EPM**
   - Predict log-running time [Hutter et al. AIJ 2014]
   - Impute right-censored data [Schmee & Hahn 1979; Hutter et al. 2011]
3. **Run efficient ablation analysis**

AClib Benchmarks

Benchmark	#P	κ [sec]	#Inst. Train/Test	Budget [h]	#Data	
SPEAR-QCP	26	5	976/2000	80	$200k$	SAT
SPEAR-SWV	26	300	302/302	768	$200k$	SAT
CPLEX-RCW2	76	10 000	495/495	768	$33k$	MIP
CLASP-WS	99	900	240/240	1536	$119k$	ASP
LPG-SATELLITE	66	300	2000/2000	768	$200k$	Planning
LPG-DEPOTS	66	300	2000/2000	768	$200k$	Planning

→ **Broadly applicable!**

see www.aclib.net

Expected Penalized Runtime

Distribution of running time prediction:

$$\int_0^{\kappa} t \cdot p(t)\mathrm{d}t + \int_{\kappa}^{\infty} X \cdot \kappa \cdot p(t)\mathrm{d}t$$

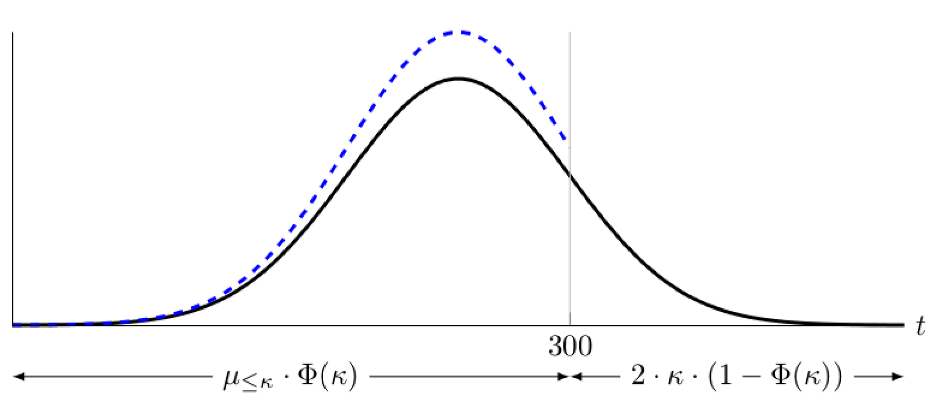
How to approximate?

- mean $\mu_{\leq\kappa}$ from truncated normal distribution $\mathcal{N}(\mu, \sigma^2)_{\leq\kappa}$ with μ being predicted running time and σ^2 predicted variance
- $\Phi(\cdot)$ is CDF of $\mathcal{N}(\mu, \sigma^2)$

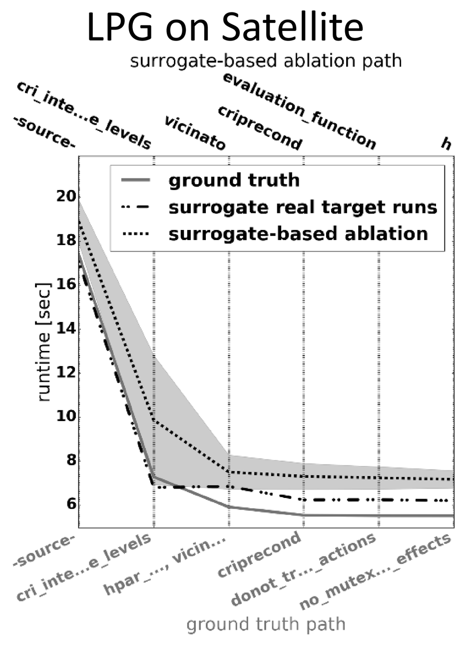
$$\mu_{\leq\kappa} \cdot \Phi(\kappa) + X \cdot \kappa \cdot (1 - \Phi(\kappa))$$

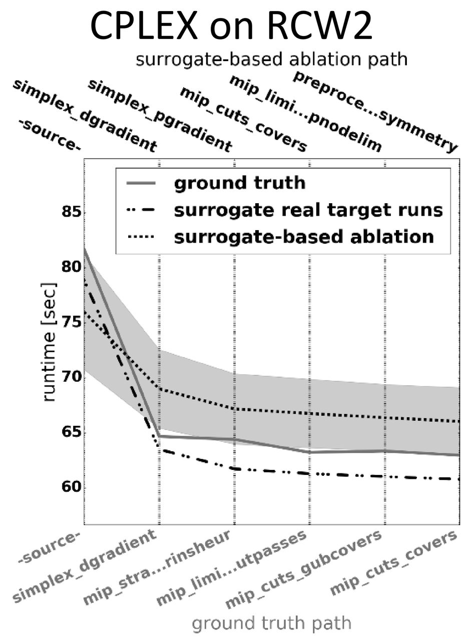
Example:

- PAR2 (X=2) with $\kappa = 300$, and predictions $\mu = 250$ and $\sigma = 50$
- Black is $\mathcal{N}(\mu, \sigma^2)$; dashed blue is $\mathcal{N}(\mu, \sigma^2)_{\leq\kappa}$
- Expected PAR2 is 293



Example Ablation Paths





Running Time [min]

benchmark	Full Train	Full Test	Racing Train	Racing Test	Surro. Train	Surro. Test
SPEAR-QCP	921	78	91	68	4	0.75
SPEAR-SWV	853	44	316	71	1	0.20
CPLEX-RCW2	121 279	11 639	21 290	11 552	2	0.23
CLASP-WS	159 799	8 266	57 689	8 323	6	0.50
LPG-DEPOTS	30 556	1 023	366	1 060	10	0.80
LPG-SATELLITE	113 126	6 533	5 783	6 162	21	1.17

Train: compute ablation path   Test: validate on test instances

Available at: http://www.ml4aad.org