Self-Optimized Strategy for IO Accelerator Parametrization

Lionel Vincent, Gaël Goret, Mamady Nabe, Trong Ton Pham

Bull - Atos Technologies
We are THE European IT Leader and a top 5 Digital services player worldwide

€ 12bn in 2016

100,000 Business technologists

72 countries
This is our Mission within Atos

1 Mission, 3 Go-to-Markets, 3 Business Units

WorldLine

Business & Platforms Solutions

Big Data & Security
700M€, 3500 people

Infrastructure & Data Management

BIG DATA & HPC
High-performance technologies for insight platforms & data management

CYBER SECURITY
Intelligent platforms for digital security, trust & compliance

MISSION-CRITICAL SYSTEMS
Intelligent systems for Defence, Homeland Security, Aerospace & Transport

Intelligence in a data-driven world
Big Data & HPC Business Unit
End-to-End offering to handle the most complex challenges

Supercomputers

Data Management

Software

HPCaaS/DLaaS

Data center

Expertise & services
BDS - HPC R&D - Data Management

Product Overview

Fast IO Libraries & Smart Burst Buffer

Application

IO Instrumentation

IO Pattern Analyzer

- Bull IO Instrumentation V1.1
- Bull Fast IO Libraries V1.0
- Bull Smart Burst Buffer (End 2018)
- Bull IO Pattern Analyzer
  - Compare jobs via GUI
  - Estimate accelerability
  - Automate FIOL activation
  - Automate accelerators parametrization (In dev.)

5s timeframe
Bull IO Pattern Analyzer

Automate accelerators parametrization
Bull IO Pattern Analyzer
Targeted User Feature

- Run a job with the minimal execution time

How to find the **optimal accelerators parametrization**?

Job Submission

IO metrics feedback
Parametrization approaches

- Static parametrization
  - Mean performance for a wide range of applications: never optimal

- Dynamic parametrization
  - Online analysis of application behavior: too costly (high overhead)

- Self-optimized parametrization
  - Find the optimal static parametrization for the job: trade-off static/dynamic
Bull IO Pattern Analyzer (IOPA)
Targeted User Feature

- Run a job with the minimal execution time
  - **Automatically** find the optimal parametrization of the IO accelerators

![Diagram showing the process of job submission and IO metrics feedback.](image)
Self-optimized parametrization method

IO Instrumentation

IO Accelerators

Cluster

IO Pattern Analyzer

Inference

Relevant runs
- Parameters
- Perf Value

Threshold

Discriminator

Topoogy

Parameters

Free Meta-data
(To be optimized)

Family Filter

Sub-set of family related runs

Runs Database
- Metrics
- Parameters
- Perf Value

Free Meta-data
(User defined)

Fixed Meta-data

Meta-data

Relevant runs

• Parameters
• Perf Value

Clustering
(Outlier detection)

Threshold

Proposed Meta-data

All runs
Inference
Inference
The « equation »

- Inference (of optimal Parameters)
- Modeling the objective function
- Numerical optimization

- Theoretical
- Interpolation
- Regression
Inference - Regression
From principle to application

Find a model which estimates the relationship:
\[ f(\text{param}) = \text{perf} \]

Different methods studied
- Bayesian Ridge Regression (BRR)
- Kernel Ridge Regression (KRR)
- Support Vector Machines for Regression (SVR)
- Gaussian Process Regression (GPR)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Prediction time*</th>
<th>Train time*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVR</td>
<td>0,001 s</td>
<td>0,110 s</td>
</tr>
<tr>
<td>KRR</td>
<td>0,001 s</td>
<td>0,120 s</td>
</tr>
<tr>
<td>GPR</td>
<td>0,001 s</td>
<td>0,069 s</td>
</tr>
</tbody>
</table>

* Time measured to perform regression/prediction on 168 runs
Inference - Optimization

Gradient-free optimization methods

- Find the optimal solutions of the IO accelerator parameters (min execution time)
  - Use heuristics to find « good » solutions in a reasonable time
- Gradient-free method
  - Less sensitive to local minimum locking

Particle Swarm Optimization (PSO)

- Use collective behavior to model the problem
- They are described by a position and a velocity

Nelder-Mead (NM)

- A simplex inspired method
- Based on four main transformations

Covariance Matrix Adaptation Evolution Strategy (CMA-ES)
Inference
Convergence validation: simulations

To be validated on real jobs

on-going work
Inference
Perspectives

- Chose the most relevant optimization algorithm

- Setup a parametrization strategy for initialization

- Implement optimization and regression features into our IOPA product
Thank you