

Machine Learning Based Automatic Performance Analysis

Flavio Alles Rodrigues Lucas Mello Schnorr

`{farodrigues,schnorr}@inf.ufrgs.br`



XV Escola Regional de Alto Desempenho
April 22nd-24th, 2015

Outline

1 Overview

How to Perform the Analysis?

Issues with the Most Common Performance Analysis Techniques

Which is the Best Way Forward?

2 Proposal

How Do We Overcome Automatic Analysis' Main Issue?

Which is the Best Way Forward?

Performance Analysis via Data Aggregation

Unsupervised Learning

What Can Be Explored?

Data, Tools & Applications

Outline

1 Overview

How to Perform the Analysis?

Issues with the Most Common Performance Analysis Techniques

Which is the Best Way Forward?

2 Proposal

How Do We Overcome Automatic Analysis' Main Issue?

Which is the Best Way Forward?

Performance Analysis via Data Aggregation

Unsupervised Learning

What Can Be Explored?

Data, Tools & Applications

Overview

How to Perform the Analysis?

Overview

How to Perform the Analysis?

- ▶ Profiling
 - ▶ A statistical summary of the observed events [Shende et al. 2006]

Overview

How to Perform the Analysis?

- ▶ Profiling
 - ▶ A statistical summary of the observed events [Shende et al. 2006]
- ▶ Visualization-based Techniques
 - ▶ Behavioral [Pagano et al. 2014]
 - ▶ Structural [Schnorr, Legrand, et al. 2013]

Overview

How to Perform the Analysis?

- ▶ Profiling
 - ▶ A statistical summary of the observed events [Shende et al. 2006]
- ▶ Visualization-based Techniques
 - ▶ Behavioral [Pagano et al. 2014]
 - ▶ Structural [Schnorr, Legrand, et al. 2013]
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Transform the data and then perform the analysis over the transformed data through a visualization-based technique [Schnorr, Huard, et al. 2012]

Overview

How to Perform the Analysis?

- ▶ Profiling
 - ▶ A statistical summary of the observed events [Shende et al. 2006]
- ▶ Visualization-based Techniques
 - ▶ Behavioral [Pagano et al. 2014]
 - ▶ Structural [Schnorr, Legrand, et al. 2013]
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Transform the data and then perform the analysis over the transformed data through a visualization-based technique [Schnorr, Huard, et al. 2012]
- ▶ Automatic Analysis
 - ▶ Detection of behavioral patterns and performance issues in a completely automated fashion [Gerndt et al. 2010][Geimer, Markus et. al. 2010]

Overview

Issues with the Most Common Performance Analysis Techniques

Overview

Issues with the Most Common Performance Analysis Techniques

- ▶ Profiling
 - ▶ Poor Data

Overview

Issues with the Most Common Performance Analysis Techniques

- ▶ Profiling
 - ▶ Poor Data
- ▶ Visualization-based Techniques
 - ▶ Poor Scalability

Overview

Issues with the Most Common Performance Analysis Techniques

- ▶ Profiling
 - ▶ Poor Data
- ▶ Visualization-based Techniques
 - ▶ Poor Scalability
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Loss of Information

Overview

Issues with the Most Common Performance Analysis Techniques

- ▶ Profiling
 - ▶ Poor Data
- ▶ Visualization-based Techniques
 - ▶ Poor Scalability
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Loss of Information
- ▶ Automatic Analysis
 - ▶ Detection of performance issues limited by known patterns

Overview

Which is the Best Way Forward?

- ▶ Profiling
 - ▶ Poor Data
- ▶ Visualization-based Techniques
 - ▶ Poor Scalability
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Loss of Information
- ▶ Automatic Analysis
 - ▶ Detection of performance issues limited by known patterns

Overview

Which is the Best Way Forward?

- ▶ Profiling
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ Data Aggregation/Transformation Followed by Visual Analysis
 - ▶ Loss of Information
- ▶ Automatic Analysis
 - ▶ Detection of performance issues limited by known patterns

Overview

Which is the Best Way Forward?

- ▶ ~~Profiling~~
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ **Data Aggregation/Transformation Followed by Visual Analysis**
 - ▶ Loss of Information
- ▶ **Automatic Analysis**
 - ▶ Detection of performance issues limited by known patterns

Overview

Which is the Best Way Forward?

- ▶ Profiling
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ ~~Data Aggregation/Transformation Followed by Visual Analysis~~
 - ▶ Loss of Information
- ▶ **Automatic Analysis**
 - ▶ Detection of performance issues limited by known patterns

Outline

1 Overview

How to Perform the Analysis?

Issues with the Most Common Performance Analysis Techniques

Which is the Best Way Forward?

2 Proposal

How Do We Overcome Automatic Analysis' Main Issue?

Which is the Best Way Forward?

Performance Analysis via Data Aggregation

Unsupervised Learning

What Can Be Explored?

Data, Tools & Applications

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ Classification & Regression (Supervised Learning)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ Classification & Regression (Supervised Learning)
 - ▶ Clustering & Outlier Detection (Unsupervised Learning)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ Classification & Regression (Supervised Learning)
 - ▶ Clustering & Outlier Detection (Unsupervised Learning)
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ **Classification & Regression (Supervised Learning)**
 - ▶ Clustering & Outlier Detection (Unsupervised Learning)
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ Clustering & Outlier Detection (Unsupervised Learning)
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ **Clustering & Outlier Detection (Unsupervised Learning)**
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ ~~Clustering & Outlier Detection (Unsupervised Learning)~~
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ ~~Clustering & Outlier Detection (Unsupervised Learning)~~
 - ▶ Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)

Proposal

How Do We Overcome Automatic Analysis' Main Issue?

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ ~~Clustering & Outlier Detection (Unsupervised Learning)~~
 - ▶ ~~Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)~~

Proposal

Which is the Best Way Forward?

- ▶ Profiling
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ ~~Data Aggregation/Transformation Followed by Visual Analysis~~
 - ▶ Loss of Information
- ▶ **Automatic Analysis**
 - ▶ Detection of performance issues limited by known patterns

Proposal

Which is the Best Way Forward?

- ▶ Profiling
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ ~~Data Aggregation/Transformation Followed by Visual Analysis~~
 - ▶ Loss of Information
- ▶ **Automatic Analysis**
 - ▶ Detection of performance issues limited by known patterns

Proposal

Which is the Best Way Forward?

- ▶ ~~Profiling~~
 - ▶ Poor Data
- ▶ ~~Visualization-based Techniques~~
 - ▶ Poor Scalability
- ▶ **Data Aggregation/Transformation Followed by Visual Analysis**
 - ▶ Loss of Information
- ▶ ~~Automatic Analysis~~
 - ▶ Detection of performance issues limited by known patterns

Proposal

Performance Analysis via Data Aggregation

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ ~~Clustering & Outlier Detection (Unsupervised Learning)~~
 - ▶ ~~Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)~~

Proposal

Performance Analysis via Data Aggregation

- ▶ Machine Learning
 - ▶ Set of Techniques that Enable the Machine to Extract from Data the Algorithm for the Task Being Computed [Alpaydin 2014]
 - ▶ ~~Classification & Regression (Supervised Learning)~~
 - ▶ **Clustering & Outlier Detection (Unsupervised Learning)**
 - ▶ ~~Reinforcement Learning (i.e. How Software Agents Ought to Behave in an Environment)~~

Proposal

Unsupervised Learning

Proposal

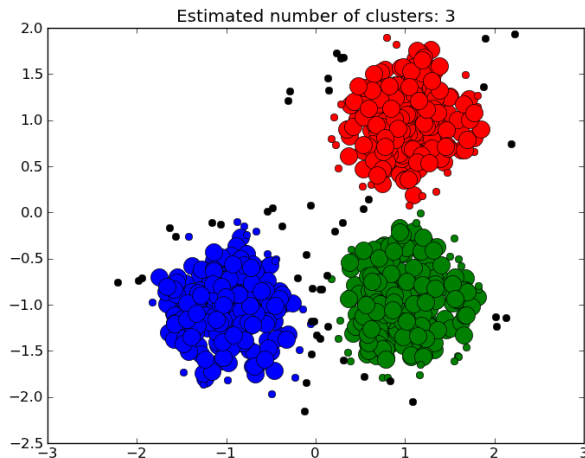
Unsupervised Learning

- ▶ The task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups

Proposal

Unsupervised Learning

- The task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups



Proposal

What Can Be Explored?

Proposal

What Can Be Explored?

- ▶ Algorithms
 - ▶ Representative-based Clustering (e.g. K-Means [MacQueen et al. 1967])
 - ▶ Density-based Clustering (e.g. DBSCAN [Ester et al. 1996])

Proposal

What Can Be Explored?

- ▶ Algorithms
 - ▶ Representative-based Clustering (e.g. K-Means [MacQueen et al. 1967])
 - ▶ Density-based Clustering (e.g. DBSCAN [Ester et al. 1996])
- ▶ Similarity Measures

Proposal

Data, Tools & Applications

- ▶ Traces
- ▶ libpaje¹
- ▶ Task-based Runtimes [Augonnet et al. 2011]

¹PajeNG on Github

References

-  Alpaydin, Ethem (2014). *Introduction to machine learning*. MIT press.
-  Augonnet, Cédric et al. (2011). “StarPU: a unified platform for task scheduling on heterogeneous multicore architectures”. In: *Concurrency and Computation: Practice and Experience* 23.2, pp. 187–198.
-  Ester, Martin et al. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *Kdd*. Vol. 96. 34, pp. 226–231.
-  Geimer, Markus et. al. (2010). “The Scalasca performance toolset architecture”. In: *Conc. & Comp.: Prac. Exp.* 22.6, pp. 702–719.
-  Gerndt, Michael and Michael Ott (2010). “Automatic performance analysis with periscope”. In: *Concurrency and Computation: Practice and Experience* 22.6, pp. 736–748.
-  MacQueen, James et al. (1967). “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA., pp. 281–297.

References

-  Pagano, Generoso and Vania Marangozova-Martin (2014). “The frameSoC software architecture for multiple-view trace data analysis”. In: *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, pp. 217–222.
-  Schnorr, Lucas, Guillaume Huard, and Philippe Navaux (2012). “A hierarchical aggregation model to achieve visualization scalability in the analysis of parallel applications”. In: *Parallel Computing* 38.3, pp. 91–110.
-  Schnorr, Lucas, Arnaud Legrand, and Jean-Marc Vincent (2013). “Interactive analysis of large distributed systems with scalable topology-based visualization”. In: *Performance Analysis of Systems and Software (ISPASS), 2013 IEEE International Symposium on*, pp. 64–73.
-  Shende, Sameer and Allen Malony (2006). “The TAU parallel performance system”. In: *International Journal of High Performance Computing Applications* 20.2, pp. 287–311.

Machine Learning Based Automatic Performance Analysis

Flavio Alles Rodrigues Lucas Mello Schnorr

`{farodrigues,schnorr}@inf.ufrgs.br`



XV Escola Regional de Alto Desempenho
April 22nd-24th, 2015