Università della Svizzera italiana

Faculty of Informatics

Data Analysis for Performance Annotations

Student: Irene Jacob **Advisor:** Prof. Antonio Carzaniga **Assistant:** Daniele Rogora

Abstract

Identifying and resolving performance issues in large software systems can be a difficult task. In particular, it is difficult to know to what extent a piece of code affects the overall performance of a system by simply analyzing the source code. It is even harder to understand why this happens. In order to support software engineers in these tasks, we want to gain more information about the performance characteristics of a piece of code. In traditional performance profiling, one measures and associates aggregate performance metrics (such as execution time) with various parts of the code (e.g. methods). In this project we aim to go further, by relating performance metrics of specific executions with relevant features of the input or the state of the system within those executions (for example, relating the size of the input with the execution time). We then perform statistical analysis of the performance and feature data, which might lead to further analysis of the code and possibly more indicative features. The ultimate goal of this analysis is to formulate synthetic performance annotations for the system and its components.

Case Studies

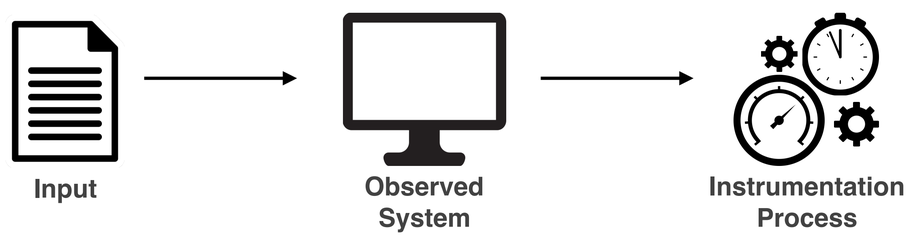
H2

Lucene

SunFlow RENDER SYSTEM The Open Source Render Engine

Feature Types

1. Value of integer parameter
2. Length of string parameter
3. Array size
4. Collection size
5. Unknown features
   - 5.1. Receiver object
   - 5.2. Object parameter particular to the scope or system being instrumented



scope: com.google.javascript.jscomp.CodeGenerator.unrollBinaryOperator

Argument Position	Feature Type	Feature Value	Performance Metric Value
2	String length	15	439346
5	Integer value	9	439346
6	Integer value	10	439346
2	String length	12	101358
5	Integer value	12	101358
6	Integer value	13	101358
2	String length	10	301445
5	Integer value	0	301445
6	Integer value	0	301445
2	String length	15	791949

An example of feature-value pairs

scope: com.google.javascript.jscomp..**

Method Name	PCC Value
buildAnnotationNames	0.99
stopTracer	0.79
reportLineCut	-0.62
parseSyntheticCode	0.98
findJsFiles	0.86

PCC values for some methods in chosen scope

Feature-Value Pairs

Correlation Analysis

Visual Representation

Further Analysis

Search for popular methods

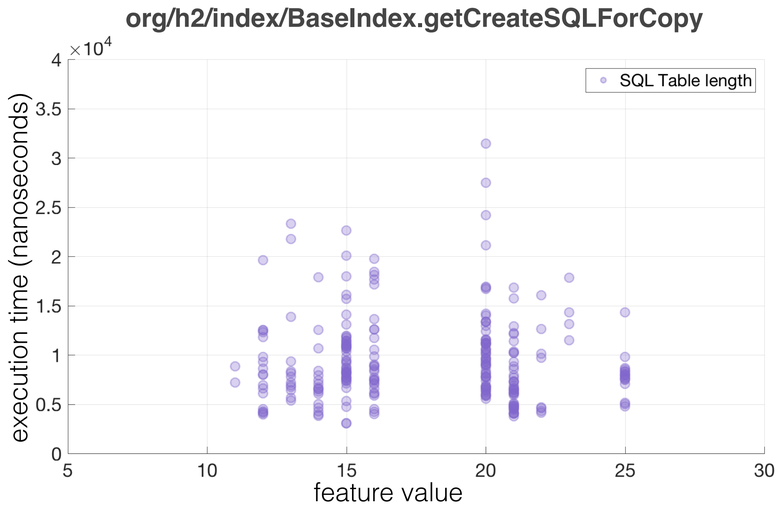
Find interesting scopes

Instrument methods in chosen scope

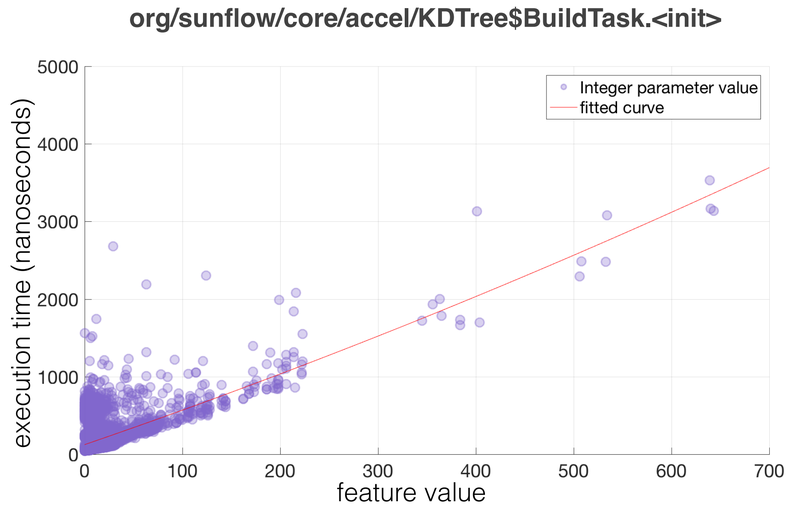
PCC

The Pearson correlation coefficient is used to measure the linear correlation between the value of the feature (e.g. array size) and the performance metric (execution time). A value close to 1 shows a strong positive correlation, a value close to -1 shows a strong negative correlation and a value near 0 shows no linear correlation.

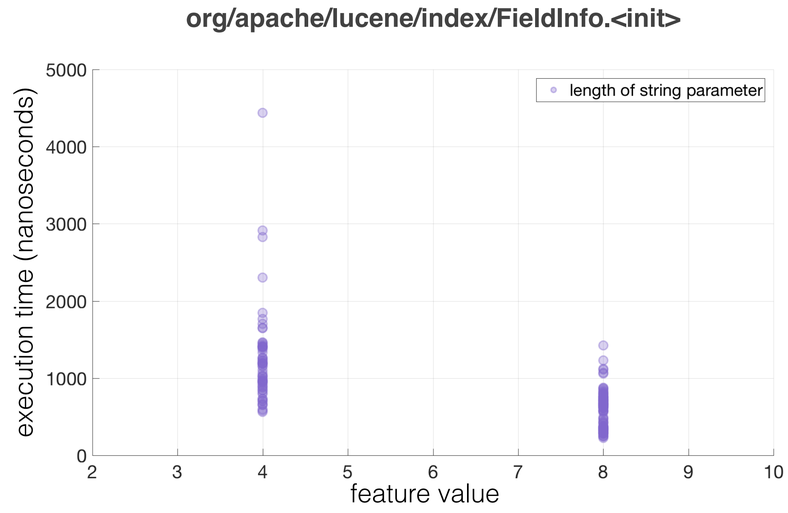
org/h2/index/BaseIndex.getCreateSQLForCopy



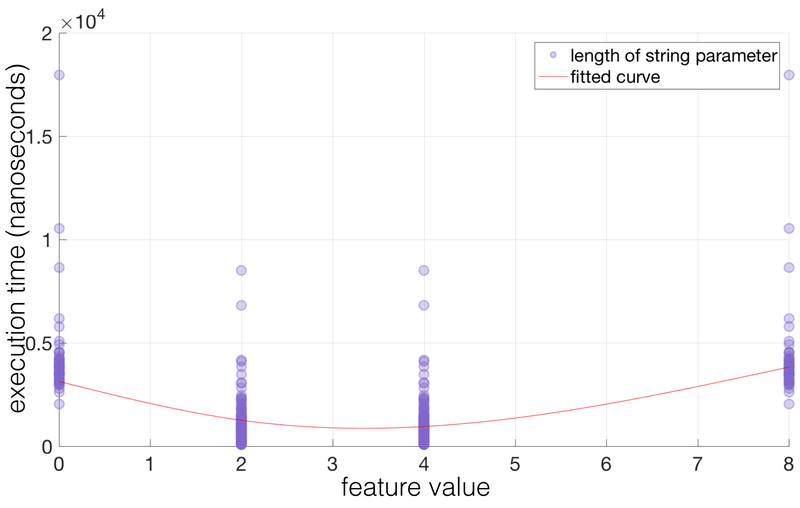
org/sunflow/core/accel/KDTree$BuildTask.<init>



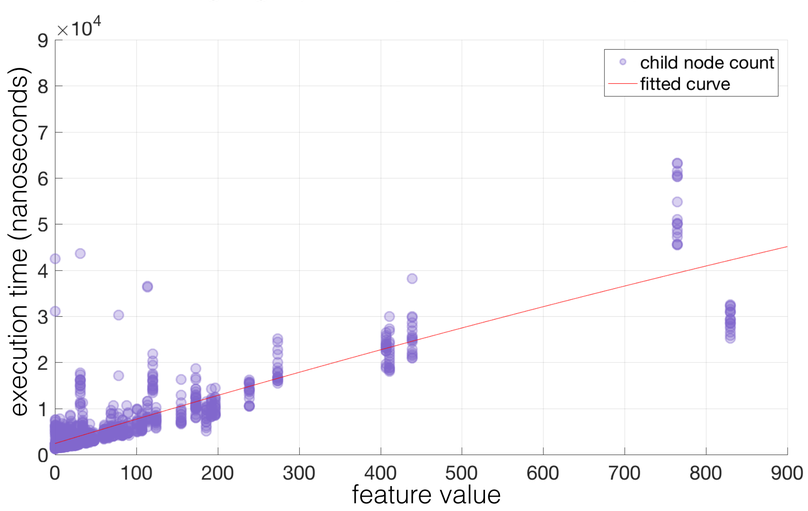
org/apache/lucene/index/FieldInfo.<init>



org/apache/lucene/index/IndexFileNames.fileNameFromGeneration



com/google/javascript/rhino/Node.setInputId



com/google/javascript/jscomp/NodeTraversal.traverse