Comparison of Analytical and Empirical Performance Models

A Case Study on Multigrid Systems

Christian Kaltenecker
University of Passau, Germany
Team

Christian Kaltenecker

Alexander Grebhahn

Sven Apel
Configurable Systems

Which configuration is performance-optimal?
# Performance Models

<table>
<thead>
<tr>
<th>Analytical</th>
<th>Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Created by domain experts</td>
<td>- Created by tools with measurement results of a specific hardware</td>
</tr>
<tr>
<td>+ Applicable on different hardware</td>
<td>+ Applicable on complex systems</td>
</tr>
<tr>
<td>- Requires domain knowledge</td>
<td>± Requires performance measurements</td>
</tr>
<tr>
<td>- Difficult for complex software</td>
<td>- Only applicable on measured hardware</td>
</tr>
</tbody>
</table>

- Mathematical formulas for estimating the performance
Performance Models

• Analytical:
  \[ 23.37 \times \log_2(px) + 23.37 \times \log_2(nx) + 2.34 \times 10^{-4} \times nx \]

• Empirical:
  \[ -5 - 3 \times px + 1.5 \times px^2 + 10^{-5} \times nx \]
Comparison Strategies

Syntactic

Compares the coefficients of the variables

\[ T_{\text{analytical}} = 2.34 \times 10^{-4} \times nx \]
\[ T_{\text{empirical}} = 10^{-5} \times nx \]

Semantic

Compares the prediction results

Hybrid

Uses both syntactic and semantic elements for the comparison

\[ T_{\text{analytical}} = 2.34 \times 10^{-4} \times nx \]
\[ T_{\text{empirical}} = 10^{-5} \times nx \]

Application of:
- Error Rate
- Jaccard Similarity
- Pearson Correlation

Calculate influence and use it as weight for Pearson correlation
Conversion of the Performance Models

- Sort each term into equivalence classes

\[ 23.37 \times \log_2(px) + 23.37 \times \log_2(nx) + 2.34 \times 10^{-4} \times nx \]

<table>
<thead>
<tr>
<th>[px]</th>
<th>[nx]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_{analytical} )</td>
<td>( 23.37 \times \log_2(px) ) ; ( 23.37 \times \log_2(nx) ) ; ( 2.34 \times 10^{-4} \times nx )</td>
</tr>
</tbody>
</table>
## Syntactic Comparison

- Use the equivalence classes and compute score according to the formula:

\[
\text{scoreOfTerms}(e, a) = \begin{cases} 
-1, & \text{if the other model has no such equivalence class} \\
0, & \text{if the equivalence class exists, but no such term} \\
1 + \text{simValue}(e, a), & \text{if the term exists in the other model}
\end{cases}
\]

\[
\text{simValue}(e, a) = \max(0, 1 - \frac{|e-a|}{\max(e, a)})
\]

\[
\begin{align*}
23.37 \times \log_2(p_x) + 23.37 \times \log_2(n_x) + 2.34 \times 10^{-4} \times n_x \\
-5 - 3 \times p_x + 10^{-5} \times n_x + 1.5 \times p_x^2
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>[constant]</th>
<th>[px]</th>
<th>[nx]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T_{\text{analytical}})</td>
<td></td>
<td>23.37 \times \log_2(p_x)</td>
<td>23.37 \times \log_2(n_x); 2.34 \times 10^{-4} \times n_x</td>
</tr>
<tr>
<td>(T_{\text{empirical}})</td>
<td>-5</td>
<td>-3 \times p_x; 1.5 \times p_x^2</td>
<td>10^{-5} \times n_x</td>
</tr>
<tr>
<td>Score</td>
<td>-1</td>
<td>0</td>
<td>1 + 0.43</td>
</tr>
</tbody>
</table>

Is this good or bad?

\[\text{Score} = 0.43\]
Semantic Comparison

- Compute the results of the performance models
Semantic Comparison

• Apply different measures on the computed results

<table>
<thead>
<tr>
<th>Name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Rate</td>
<td>10.1%</td>
</tr>
<tr>
<td>Jaccard Similarity</td>
<td>90%</td>
</tr>
<tr>
<td>Manhattan Distance</td>
<td>0.046</td>
</tr>
<tr>
<td>Euclidean Distance</td>
<td>0.0029</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.41 (medium)</td>
</tr>
</tbody>
</table>
Hybrid Comparison

• Calculate the influence of the equivalence classes:

<table>
<thead>
<tr>
<th></th>
<th>[constant]</th>
<th>[px]</th>
<th>[nx]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{analytical}}$</td>
<td></td>
<td>23.37 $\times$ $\log_2(px)$</td>
<td>23.37 $\times$ $\log_2(nx); 2.34 \times 10^{-4} \times nx$</td>
</tr>
<tr>
<td>Influence$_{\text{analytical}}$</td>
<td>0%</td>
<td>1.1%</td>
<td>98.9%</td>
</tr>
<tr>
<td>$T_{\text{empirical}}$</td>
<td></td>
<td>$-3 \times px; 1.5 \times px^2$</td>
<td>10$^{-5}$ $\times nx$</td>
</tr>
<tr>
<td>Influence$_{\text{empirical}}$</td>
<td>0%</td>
<td>71.2%</td>
<td>28.8%</td>
</tr>
</tbody>
</table>

• Use the similarity of the influences as a weight for the Pearson correlation between the equivalence classes
Result: 0.3
Conclusion

Performance Models

- Mathematical formulas for estimating the performance

Analytical
- Created by domain experts
- Applicable on different hardware
- Requires domain knowledge
- Difficult for complex software

Empirical
- Created by tools with measurement results of a specific hardware
- Applicable on complex systems
- Requires performance measurements
- Only applicable on measured hardware

Syntactic Comparison

- Use the equivalence classes and compute score according to the formula:

\[
\text{score}(\text{formula}, a) = \left\{ \begin{array}{ll}
\text{error} & \text{if the other model has no such equivalence class} \\
\text{error} & \text{if the equivalence class exists but no such term} \\
\text{error} & \text{if the terms exist in the other model but not in this formula} \\
\text{error} & \text{if the terms exist in both models but are different}
\end{array} \right.
\]

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Score</th>
<th>nx</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>-4.3</td>
<td>1.5</td>
<td>-10.5</td>
</tr>
<tr>
<td>Score</td>
<td>0</td>
<td>0</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Hybrid Comparison

- Calculate the influence of the equivalence classes:

<table>
<thead>
<tr>
<th>Syntax</th>
<th>constant</th>
<th>px</th>
<th>nx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>23.37 + log(px)</td>
<td>23.37 + log(nx); 2.34 + 10^{-4} * nx</td>
<td></td>
</tr>
<tr>
<td>Influence Target</td>
<td>0%</td>
<td>27%</td>
<td>27%</td>
</tr>
<tr>
<td>Target</td>
<td>-5</td>
<td>-3 + px + 1.5 * px^2</td>
<td>18 + nx</td>
</tr>
<tr>
<td>Influence Target</td>
<td>0%</td>
<td>27%</td>
<td>27%</td>
</tr>
</tbody>
</table>

- Use the influence as a weight for the Pearson correlation between the equivalence classes
Result: 0.15 (weak)
Thank you for your attention!