Simulation Data Mining for Supporting Bridge Design

Ninth Australasian Data Mining Conference

Steven Burrows†  Benno Stein†  Jörg Frochte‡
David Wiesner†  Katja Müller†

† Bauhaus University Weimar
‡ Bochum University of Applied Science

1–2 December 2011
About Me

- Undergrad, Honours, PhD, RMIT University, up to 2010.
- PostDoc, Bauhaus University Weimar, 2011 to current.
- “Strategies for Robust Design of Structures” project.
  - Includes simulation data mining and civil engineering sub-projects.
Applications of Simulation Data Mining

- Car crashworthiness (Kuhlmann et al., 2005; Mei and Thole, 2007).
- Occupant restraint systems (Zhao et al., 2010).
- Aviation (Fayyad et al., 1996; Painter et al., 2006).
- Semiconductor manufacturing (Brady and Yellig, 2005).

Interactive Bridge Design in Civil Engineering

Parameterization → Inference → Evaluation → Simulation

Model in the design space

Analysis result in the simulation space (overlayed over design space)
Supporting Bridge Design

Key idea:
- Mine patterns in pre-computed bridge simulation results.

Why simulation data mining?:
- Faster simulations, provide diagnosis, automated design, etc.

Consider models \( \{ m_i \in M \} \) and simulation results \( \{ y_i \in Y \} \):
- \( \| y_1 \ominus y_2 \| < \varepsilon \iff \varphi_{\text{Design}}(m_1, m_2) \approx 1. \)
- Develop \( \varphi_{\text{Design}} \) to predict the similarity of two designs with regards to learned behavior.

Questions we can answer:

1. Identify good ‘?’ in \( \varphi_{\text{Design}}(m_1, ?) \approx 1. \)
2. Predict the behavior of a new model \( m \).
3. Learn cost optimization rules for any equivalence class \( M' \subseteq M \).
Methodology

Computation of the similarity measure in six steps:

1. Sample candidate designs.
2. Simulate the models.
3. Aggregate the simulation results.
4. Cluster the simulation results.
5. Sample the simulation results.
6. Learn a mapping from \( \{m_i \in M\} \) to \( \{y_i \in Y\} \).

Future work disclaimer:
- There are still competing alternatives in many steps to be explored.
Step 1: Sample Candidate Designs

Data format:
- IFC (Industry Foundation Classes): An object-oriented data model for describing entities in the construction and building industries.
- IFC-Bridge: An extension to IFC for bridges.
- NURBS (Non-Uniform Rational Basis Spline): Novel extension to IFC-Bridge in the project.

Data set:
- 14,641 geometry and material permutations of the model below.
Step 2: Simulate the Models

Input:
- IFC-Bridge data models (Lebegue et al., 2007) with NURBS.

Simulation Engine:
- Finite Element Method implementation (Gerold, 2010). Our “oracle”.

Output:
- VTK (Visualization Toolkit) format (Schroeder et al., 1996).
Step 3: Aggregate the Simulation Results

Original data:
- 12,064 points and measurements from the FEM mesh.

Process:
- Consultation with a Numerics professor.

Aggregated data (45 measurements):
- Five regions (below).
- Maximum displacement, strain, and stress.
- X, Y, and Z co-ordinates.
Step 4: Cluster the Simulation Results

Goal:
- Learn similar groupings of simulated models.

Clustering algorithms:
- AiTools implementation (http://webis.de/research/projects/aitools).

Evaluation:
- Expected Density measure (Stein et al., 2003).
- Higher quality clusterings give have higher expected density score.
Step 5: Sample the Simulation Results

Example (350 items):

Cluster A: 100 items. $= 4950 \left( \frac{100(100-1)}{2} \right)$ positive pairs.
Cluster B: 120 items. $= 7140 \left( \frac{120(120-1)}{2} \right)$ positive pairs.
Cluster C: 130 items. $= 8385 \left( \frac{130(130-1)}{2} \right)$ positive pairs.

= 20475 positive pairs.
+ 40600 negative pairs.

= 61075 total pairs.

Sampling strategy (from approximately $10^8$ pairs):

- Class balance.
- Equal sampling from each cluster of positive pairs.
- Random sampling for negative pairs.
Step 6: Machine Learning

Training data:
- Duples in the form \( \langle m_k \ominus m_l, c_j \rangle \).

Learning:
- Ten-fold cross validation.
- Naive Bayes and Maximum Entropy classifiers (Burrows et al., 2011).

Outcome:
- Class probability estimates \([0, 1]\) for evaluating \( \varphi_{\text{Design}} \).
Clustering Results

![Graph showing clustering results with expected density on the y-axis and number of clusters per clustering on the x-axis. The graph compares k-means (green solid line) and HAC (teal dashed line).]
## Accuracy Results

<table>
<thead>
<tr>
<th>Data set size</th>
<th>Naive bayes</th>
<th>Entropy</th>
<th>Naive bayes</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>94.0</td>
<td>94.0</td>
<td>94.0</td>
<td>96.0</td>
</tr>
<tr>
<td>200</td>
<td>92.5</td>
<td>93.0</td>
<td>90.0</td>
<td>91.0</td>
</tr>
<tr>
<td>500</td>
<td>85.4</td>
<td>90.6</td>
<td>89.8</td>
<td>90.8</td>
</tr>
<tr>
<td>1,000</td>
<td>91.4</td>
<td>94.3</td>
<td>90.8</td>
<td>91.2</td>
</tr>
<tr>
<td>2,000</td>
<td>88.6</td>
<td>92.4</td>
<td>89.8</td>
<td>91.0</td>
</tr>
<tr>
<td>5,000</td>
<td>89.4</td>
<td>92.7</td>
<td>89.3</td>
<td>90.5</td>
</tr>
<tr>
<td>10,000</td>
<td>89.6</td>
<td>92.4</td>
<td>88.5</td>
<td>89.4</td>
</tr>
<tr>
<td>20,000</td>
<td>89.8</td>
<td>92.3</td>
<td>89.6</td>
<td>90.3</td>
</tr>
<tr>
<td>50,000</td>
<td>89.9</td>
<td>92.7</td>
<td>89.1</td>
<td>90.1</td>
</tr>
<tr>
<td>100,000</td>
<td>89.7</td>
<td>92.4</td>
<td>89.1</td>
<td>89.7</td>
</tr>
<tr>
<td>200,000</td>
<td>89.8</td>
<td>92.5</td>
<td>89.1</td>
<td>89.8</td>
</tr>
<tr>
<td>all</td>
<td>89.8</td>
<td>92.5</td>
<td>89.1</td>
<td>89.7</td>
</tr>
</tbody>
</table>
Future Work

- Use of a rank correlation co-efficient such as Spearman’s \textit{rho}, Pearson’s \textit{r}, or Kendall’s \textit{tau} to compare the correlation of the ranks of $\varphi_{\text{Design}}$ with the ranks of the cosine similarity taken from the simulation space.

- Apply clustering instead of ranking for the evaluation, and compare the coverage of the clusterings (F-measure).

- Apply \textit{domain decomposition} as a parallelization technique for solving partial differential equations in FEM analysis.
Summary

- Mine patterns in pre-computed bridge simulation results for knowledge discovery.
- Six step methodology for computing $\varphi_{Design}$ so that new questions can be answered.
- Initial results are promising, but more remains for future work.

Thankyou!

Steven Burrows
steven.burrows@uni-weimar.de
www.webis.de