IDLE: Interactive Descriptions for Low-dimensional Embedding

Aindrila Basak¹, Mona Nashaat¹, James Miller¹, Shaikh Quader²
¹University of Alberta, Edmonton, Canada, ²IBM Canada, Toronto, Canada

Research Objectives

Interpret Black-Box Models
- Most big data analysis algorithms are black-box.
- Users do not use what they don’t understand or trust.
- Leading to the loss of millions of dollars invested in building models.

Engage Users in Decision Making
- Necessity for Human-in-the-loop.
- Need to incorporate user expertise in decisions.
- Need for proactive guidance with complex data analysis.

Experimental Methodology

- Compare 15 Dimensionality Reduction Algorithms
- For 7 Evaluation Metrics
- Over 30 Real-world Datasets
- Evaluate with 6 Statistical Significance Tests

Experimental Outcome

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>Best Performers</th>
<th>Mediocre Performers</th>
<th>Worst Performers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML Accuracy</td>
<td>KernelPCA, PCA</td>
<td>Fit-SNE, LEM</td>
<td>LTSA, HLLE</td>
</tr>
<tr>
<td>Execution Time</td>
<td>PCA, Isomap</td>
<td>openTSNE, LTSA</td>
<td>MDS, LEM</td>
</tr>
<tr>
<td>Local Structure</td>
<td>MDS, openTSNE</td>
<td>Fit-SNE, UMAP</td>
<td>LLE, Isomap</td>
</tr>
<tr>
<td>Global Structure</td>
<td>MDS, KernelPCA</td>
<td>LEM, HLLE</td>
<td>Trimap, t-SNE</td>
</tr>
<tr>
<td>Outlier Effects</td>
<td>LTSA, Isomap</td>
<td>t-SNE, openTSNE</td>
<td>LLE, MLLE</td>
</tr>
<tr>
<td>Duplicate Effects</td>
<td>t-SNE, Trimap</td>
<td>HLLE, LEM</td>
<td>MDS, KernelPCA</td>
</tr>
<tr>
<td>Partial Records</td>
<td>PCA, KernelPCA</td>
<td>UMAP, Trimap</td>
<td>Fit-SNE, t-SNE</td>
</tr>
</tbody>
</table>

IDLE - Process Flow

Results

Preservation of Global Structure:
Measure of pair-wise geodesic distances among data-points in the high-dimensional manifold.