(1) Motivation

- How do developers split their time between different components of the workflow?
- Developers iteratively modify machine learning workflows using results from previous versions.

Current studies suffer from response bias.

- We conduct a statistical study of machine learning model development from empirical evidence.
- Study results reveal desiderata for ML systems.

(2.1) Data

- 105 applied data science papers published in 2016

<table>
<thead>
<tr>
<th>Domain</th>
<th>Web App</th>
<th>Social Sci</th>
<th>NLP</th>
<th>Natural Sci</th>
<th>Vision</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDD</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ACL</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>NATURE</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>CVPR</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Application domain of papers breakdown by conference

- Limitations of data
  - Incomplete picture of iterations
  - Results not presented in iteration order
  - Small per domain corpus can lead to spurious results

- Remedies
  - Multiple surveys to reduce chance of spurious results
  - Design iteration estimators that do not rely on order

(2.2) Estimating Iterations

<table>
<thead>
<tr>
<th>Data Prep.</th>
<th>ML Model Class</th>
<th>ML Tuning</th>
<th>Evaluation Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm</td>
<td>SVM</td>
<td>Reg.</td>
<td>AUC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td># tables</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td># figs</td>
</tr>
<tr>
<td>N = 2</td>
<td>N = M - 1</td>
<td>N = 2</td>
<td>N = M - 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = M - 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 2</td>
</tr>
</tbody>
</table>

- Three surveys independently identify operations mentioned in a paper.
- Take majority vote to aggregate the three results.
- Collect statistics on the aggregate annotation.

Estimator for # data prep. iterations $\hat{t}_{DPR} = n_D$.

Estimator for # ML iterations $\hat{t}_{M} = (n_M - 1) + (n_P - 1)$.

Estimator for # post proc. iter. $\hat{t}_{PPR} = \min(n_E, n_{table} + n_{figure})$.

(3.1) Common Operations

- Data prep.:
  - Human def. Features (long live feat. eng.)
  - Data join

- Model class:
  - GLM
  - SVM
  - DNN only in NLP/CV

- Model Tuning:
  - Learning rate
  - Batch size
  - Fine-grained: P/R, Accuracy

- Evaluation:
  - Coarse-grained: Case studies, Vis.

Evaluate

Prioritize system support for common operations

Open source dataset at https://github.com/helix-ml/AppliedMLSurvey