Towards understanding learning behavior

Joaquin Vanschoren

Meta-learning

How to select the right learning algorithm for a given dataset?

**Base-Learning**

1. Define instance space
2. Measure instances (build dataset)
3. Preprocess data (sampling, discretizing, feature selection,...)
4. Choose learning algorithm (+ parameter settings)
5. Predict new instances/derive rules

**Meta-Learning**

1. Define meta-features for datasets (reflecting restrictions of various learning algorithms)
2. Measure algorithm performance in experiments (build meta-dataset)
3. Preprocess experiment data
4. Choose learning algorithm (+ parameter settings)
5. Predict algorithm performance/derive rules, rankings

**“Traditional” approach**

<table>
<thead>
<tr>
<th>Dataset characteristics</th>
<th>Performance measures</th>
</tr>
</thead>
</table>
| Size | nb. Attr. | Algorithm | Acc | RunT | ...
| 2300 | 34 | C4.5 | 92 | 43 | ...

Publicly available datasets

| Id | nb. samples | ...
|---|-------------|---|
| Dataset1 | 2300 | ...

Towards Descriptive Meta-learning

**Experiment database**

| Algo Impl. | Param. sett. | Dataset | TP | FP | ...
|------------|--------------|---------|----|----|-------|
| C4.5 v.1  | C451 - 1     | Dataset1 | 945| 84 | ...

**Performance measures**

| TP | FP | ...
|----|----|---|
| 945| 84 | ...

Use of data fragmentation or attribute summations,
ability to handle fine-grained concepts or local relevance,...

**Limitations:**

1) The Curse of Dimensionality → We need many more datasets
2) Generalisation over algorithms? → Algorithms should be characterized
3) Only when, not why... → More thorough investigation needed
4) What about preprocessing? → Effects should be included

**Experiment databases**

Make experiments reusable, reproducible: log all experiment details and results

Querying or datamining on stored results

→ Thorough investigation of interactions between algorithm parameters, dataset characteristics and performance measures

**Synthetic datasets**

Unbiased: hide a large range of different kinds of concepts in the data, and
classify: model characteristics, concept variation, example cohesiveness,...

“Natural”: approximate characteristics of natural datasets: complex relations
between attributes, noise, irrelevant attributes, missing values,...

Coverage: control characteristics to cover meta-feature space: experiment design

**Algorithm characterization**

Parameter settings: parameters and techniques used (1 table/algorithm)

General algorithm properties: representation model,
dependency on linear separability or conditional independence,
use of data fragmentation or attribute summations,
ability to handle fine-grained concepts or local relevance,...

**Understanding inductive performance**

Averaged perf. measures do not explain why an algorithm failed/succeeded
→ Decompose the misclassification error: bias error vs. variance error

<table>
<thead>
<tr>
<th>dep. bias</th>
<th>comp. bias</th>
<th>bias err</th>
<th>variance err</th>
</tr>
</thead>
<tbody>
<tr>
<td>inappropriate</td>
<td>too strong</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>inappropriate</td>
<td>too weak</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>inappropriate</td>
<td>too strong</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>appropriate</td>
<td>too weak</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>

→ advice possible improvements

**Preprocessing steps**

Link dataset characteristics to effect of preprocessing techniques
→ Separate experiment database for preprocessing experiments

Characterise dataset and predict changes after preprocessing, or advice useful
preprocessing for optimizing performance of specific algorithm

Strong link with bias/variance error

Declarative Languages and Artificial Intelligence research group, Department of Computer Science, K.U.Leuven, Belgium