Station 6: Machine Learning

Michelle Cawley
mcawley@unc.edu

Health Sciences Library
University of North Carolina at Chapel Hill
Machine Learning

Deeper Dive

- Fully supervised method
- Training can be at one time or iterative
  - Active ML requires its own screening environment.
- Use cases:
  - In conjunction with other methods when training data are difficult to locate.
  - “Insurance step” to review a portion of studies excluded automatically by other methods.
Visualizing (Passive) Machine Learning

Unclassified Studies

Training Data

Machine Learning Model

Classified Studies

Predicted Relevant

Predicted Not Relevant
Visualizing Active Machine Learning

Unclassified Studies

Manually Screen High Priority Studies

Training Data

Reclassified Studies

Predicted Relevant

Predicted Not Relevant

Machine Learning Model
Supervised Clustering compared to Machine Learning

- Uses clustering algorithms
- Requires seeds
- Only positive studies are needed
- Small number of studies required

- Uses ML algorithms
- Requires training dataset
- Requires positive and negative studies
- Large volume of studies must be classified manually

Like training data, seeds are studies that have been classified manually for relevance.
Supervised Clustering with an Ensemble Approach

Options at each checkpoint:
1. Keep screening
2. Stop screening
3. Modify Approach (e.g., switch to machine learning)
**Research Question:** Does exposure to pollutants in utero lead to asthma?

<table>
<thead>
<tr>
<th>ML</th>
<th>Manual</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1. Vitamin D deficiency at 16 to 20 weeks gestation is associated with impaired lung function and asthma at 6 years of age.</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2. Characteristics of infants admitted to hospital for persistent colic, and comparison with healthy infants.</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3. Home exposures to environmental tobacco smoke and allergic symptoms among young children in Singapore.</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4. Fetal and postnatal exposure to tobacco smoke and respiratory health in children.</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5. Atopic dermatitis, extrinsic atopic dermatitis and the hygiene hypothesis: results from a cross-sectional study.</td>
<td></td>
</tr>
</tbody>
</table>
Next Steps

• Other tools available:
  • Robot Analyst | DistillerAI | Colandr | Rayyan | Abstrackr

• Keep learning
  • Look for CE Course at MLA 2020
  • Consider joining AI SIG

• Visit Station 1: Keyword Analysis Tool