The Women’s Health and Aging Study

[With Karen Bandeen-Roche]

The Women’s Health and Aging Study (WHAS) began in 1992 to study the causes and the course of disability in moderately to severely disabled older women living in the community.

The WHAS is a population-based longitudinal study of women with at least mild disability, 65 years of age or older, living at home in eastern Baltimore city or county.

There is evidence that disability results from chronic diseases, and that interactions between diseases (comorbidities) are of importance in causing disability.

In this presentation we are concerned about relating chronic diseases and their interactions to death.
Study subjects:

- 32538 women were identified by searching medicare enrollment files,
- 6521 women were sampled (age-stratified),
- 5316 women were alive and living at home,
- 4137 women participated in the home-based screening,
- 1409 women were eligible,
- 1002 women agreed to participate and provided written informed consent.

The major chronic diseases at baseline were ascertained by using complex algorithms. Follow-up evaluations were conducted every 6 months for 3 years.

The Women’s Health and Aging Study

angina  heart pain
cancer  cancer
chf  congestion heart failure
diabetes  diabetes
disc  degenerative disc disease
hf  hip fracture
mi  myocardial infarction
oatot  osteo-arthritis at hand, knee or hip
oahand  osteo-arthritis at hand
oahip  osteo-arthritis at knee
oaknee  osteo-arthritis at hip
osteo  osteoporosis
pad  peripheral arterial disease
parkin  parkinson’s disease
pulmonary  pulmonary disease
ra  rheumatoid arthritis
stenosis  spical stenosis
stroke  stroke
The Women’s Health and Aging Study

\[
p = \Pr(\text{death in round } j \mid \text{ survival to round } j-1, X, \text{ age})
\]

\[
\text{logit}(p) = -9.01 + 0.06 \cdot \text{age} + 1.07 \cdot L(X)
\]

or

and

- angina
- chf
- cancer
- diab
- stroke
Logic Regression

[With Charles Kooperberg and Michael LeBlanc]

\(X_1, \ldots, X_k\) are 0/1 (False/True) predictors.

\(Y\) is a response variable.

Fit a model

\[ g(E(Y)) = b_0 + \sum_{j=1}^{t} b_j \cdot L_j, \]

where \(L_j\) is a Boolean combination of the covariates, e.g. \(L_j = (X_1 \lor X_2) \land X_4^c\).

Determine the logic terms \(L_j\) and estimate the \(b_j\) simultaneously.

The Move Set for Logic Regression

<table>
<thead>
<tr>
<th>Possible Moves</th>
<th>Alternate Leaf</th>
<th>Alternate Operator</th>
<th>Grow Branch</th>
<th>Prune Branch</th>
<th>Split Leaf</th>
<th>Delete Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>and</td>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternate Leaf</td>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grow Branch</td>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and</td>
<td>(d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Split Leaf</td>
<td>(e)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete Leaf</td>
<td>(f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We try to fit the model \( g(E(Y)) = b_0 + \sum_{j=1}^{t} b_j \cdot L_j \).

- Select a scoring function (RSS, log-likelihood, ...).
- Pick the maximum number of Logic Trees.
- Pick the maximum number of leaves in a tree.
- Initialize the model with \( L_j = 0 \) for all \( j \).
- Carry out the Simulated Annealing Algorithm:
  - Propose a move.
  - Accept or reject the move, depending on the scores and the temperature.
number of binary predictors in the models

intercept and age
References


Software and manuscripts available at: http://biostat.jhsph.edu/~iruczins/