ExpertBayes: Automatically Refining Manually Built Bayesian Networks

ICMLA 2014 – December 4th 2014 – Detroit, USA

Ezilda Almeida
Pedro Ferreira
Tiago T. V. Vinhoza
Inês Dutra
Paulo Borges
Yirong Wu
Elizabeth Burnside
Outline

- Objectives
- Datasets
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work
Outline

• Objectives
• Datasets
• Methodology and Tools
• Results and Analysis
• ExpertBayes (graphical user interface)
• Conclusions and Future Work
Objectives

Network constructed manually

New network with better score

ExpertBayes
Outline

- Objectives
- Datasets
- Methodology and Tools
- Results and Analysis
- ExpertBayes (graphical user interface)
- Conclusions and Future Work
Dataset

- **Prostate Cancer**: 
  - 496 cases
  - Each case refers to the clinical history of each patient
- **Breast Cancer (1)**: 
  - 100 cases
  - Each case refers to a breast nodule from mammography results
- **Breast Cancer (2)**: 
  - 241 cases
  - Each case refers to a breast nodule from mammography results
Attributes

- Prostate Cancer

11 Attributes

- Age (age)
- Weight (wt)
- Family history of cancer (hx)
- Systolic blood pressure (Sbp)
- Diastolic blood pressure (Dbp)
- Hmoglobins (hg)
- Clinical stage (stage)
- Doubling time PSA (Dtime)
- Size of the prostate (size)
- Bony metastases (bm)

Status (status)

- 351 Dead (+)
- 145 Alive (-)
Attributes

- Breast Cancer(1)

33 Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Disease</td>
</tr>
<tr>
<td>BreastDensity</td>
</tr>
<tr>
<td>MassesShape</td>
</tr>
<tr>
<td>MassesDensity</td>
</tr>
<tr>
<td>MassesSize</td>
</tr>
<tr>
<td>PostOpChange</td>
</tr>
<tr>
<td>MassesStability</td>
</tr>
<tr>
<td>Calc_Milk</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td><strong>BinaryDx</strong></td>
</tr>
</tbody>
</table>

45 Benign

55 Malignant
Attributes

- Breast Cancer (2)

8 Attributes

<table>
<thead>
<tr>
<th>Age</th>
<th>Mass_Shape</th>
<th>Mass_Margins</th>
<th>Depth</th>
<th>Size</th>
<th>Overall_Breast_Composition</th>
<th>Retro_Density</th>
<th>Biopsy_Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

153 Benign

88 Malignant

(+) (−)
Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- ExpertBayes (graphical user interface)
- Conclusions and Future Work
Methodology and Tools

- Eclipse to develop ExpertBayes using Java language
- WEKA
- 5-fold cross-validation to train and test our models
- t-test was used to validate the results
  - Significance level: 0.05
Outline

• Objectives
• Dataset
• Methodology and Tools
• Results and Analysis
• *ExpertBayes* (graphical user interface)
• Conclusions and Future Work
Results and Analysis

• CCI (%) test set - across 5-folds

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Original</th>
<th>ExpertBayes</th>
<th>WEKA-K2</th>
<th>WEKA-TAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate Cancer</td>
<td>74</td>
<td>76</td>
<td>74</td>
<td>71</td>
</tr>
<tr>
<td>Breast Cancer (1)</td>
<td>49</td>
<td>63</td>
<td>59</td>
<td>57</td>
</tr>
<tr>
<td>Breast Cancer (2)</td>
<td>49</td>
<td>64</td>
<td>80</td>
<td>79</td>
</tr>
</tbody>
</table>
Results and Analysis

- Precision-Recall Curves for various thresholds
  - Prostate
Results and Analysis

• Precision-Recall Curves for various thresholds
  ▫ Breast Cancer (1)
Results and Analysis

• Precision-Recall Curves for various thresholds
  ▫ Breast Cancer (2)
Results and Analysis: prostate cancer networks

Original Network

ExpertBayes

CCI : 74%

CCI : 76%
Results and Analysis: prostate cancer networks

Weka TAN

ExpertBayes

CCI : 71%

CCI : 76%
Outline

• Objectives
• Dataset
• Methodology and Tools
• Results and Analysis
• ExpertBayes (graphical user interface)
• Conclusions and Future Work
ExpertBayes

- Graphical user interface
Outline

• Objectives
• Dataset
• Methodology and Tools
• Results and Analysis
• *ExpertBayes* (graphical user interface)
• Conclusions and Future Work
Conclusions and Future Work

- ExpertBayes produces better results than the original model and better results than models learned with other tools.

- ExpertBayes also provides a graphical user interface (GUI) where users can play with their models thus exploring new structures that give rise to a search for other models.
Conclusions and Future Work

• Improve the algorithm in order to have better prediction performance.

• Using more (and quality) data, different search and parameter learning methods.
Thank you!

ezildacv@gmail.com
pedroferreira@dcc.fc.up.pt
tiago.vinhoza@gmail.com
ines@dcc.fc.up.pt
pauloraborges@gmail.com
eburnside@uwhealth.org
Appendices
State of the Art

- Previous works considered as initial network a naive Bayes or empty network [9], [4]:


State of the Art

The R packages deal [2] and bnlearn [11], [13] can refine any input network. However, deal and bnlearn refine input networks by successive refinements instead of performing the refinement only over the original network:


State of the Art

- WEKA, whose bayesian algorithms apply successive refinements to the newly built models:
Methodology

WEKA:
• K2 is a greedy algorithm that, given an upper bound to the number of parents for a node, tries to find a set of parents that maximizes the likelihood of the class variable [6].

• TAN (Tree Augmented Naive Bayes) generates a tree over naive Bayes structure, where each node has at most two parents, being one of them the class variable [8].
## Data Distribution

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of Instances</th>
<th>Number of Variables</th>
<th>Pos.</th>
<th>Neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate Cancer</td>
<td>496</td>
<td>11</td>
<td>352</td>
<td>144</td>
</tr>
<tr>
<td>Breast Cancer (1)</td>
<td>100</td>
<td>34</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>Breast Cancer (2)</td>
<td>241</td>
<td>8</td>
<td>88</td>
<td>153</td>
</tr>
</tbody>
</table>