Predicting Disease Complications
Using a Stepwise Hidden Variable Approach
For Learning Dynamic Bayesian Networks

Leila Yousefi
Allan Tucker
Mashael Al-Luhaybi
Brunel University London

Lucia Saachi,
Riccardo Bellazzi
University of Pavia

Luca Chiovato
Instituti Maugeri
Type 2 Diabetes

What people see:
- High blood sugar

What people don’t see:
- Blindness
- Blurred vision
- Boils
- Cataracts
- Depression
- Erectile dysfunction
- Foot ulcers
- Frequent urination
- Glaucoma
- Intense fatigue
- Intense hunger
- Intense thirst
- Itchiness
- Kidney disease
- Numbness
- Pain
- Sexual dysfunction
- Skin infections
Outline

- Motivation
- Data
- Problem
- Solution
- Dynamic Bayesian Networks (DBNs)
- Hidden variable discovery approach
  - Pair sampling and Stepwise approach procedure
- Results
- Conclusions and future works

Leila.Yousefi@brunel.ac.uk
Mortality due to diabetes age 20-79 in 2017 (in millions)
The Data at Maugeri, Pavia:

- Type 2 Diabetes Mellitus (T2DM)
- Patients aged 25 to 65 years.
- IRCCS Istituti Clinici Scientifici Maugeri of Pavia, Italy.
- MOSAIC project funded by the European Commission.

- T2DM risk factors:
  - Physical examination
  - Laboratory data

- MATLAB and Bayes Net toolbox (murphy,2001)
- Visualization we used Graphviz.
<table>
<thead>
<tr>
<th>Visit NO</th>
<th>Patient ID</th>
<th>HbA1c</th>
<th>Retinopathy</th>
<th>Neuropathy</th>
<th>Nephropathy</th>
<th>Liver disease</th>
<th>Hypertension</th>
<th>BMI</th>
<th>Creatinine</th>
<th>Cholesterol</th>
<th>HDL</th>
<th>DBP</th>
<th>SBP</th>
<th>SMK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>885</td>
<td>0.769</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.286</td>
<td>-0.391</td>
<td>2.082</td>
<td>0.020</td>
<td>1.705</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>2</td>
<td>885</td>
<td>0.769</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.286</td>
<td>-0.391</td>
<td>2.082</td>
<td>0.020</td>
<td>1.705</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>3</td>
<td>885</td>
<td>0.769</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.286</td>
<td>-0.391</td>
<td>2.082</td>
<td>0.020</td>
<td>1.705</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>4</td>
<td>885</td>
<td>0.769</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.286</td>
<td>-0.391</td>
<td>2.082</td>
<td>0.020</td>
<td>1.705</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>5</td>
<td>894</td>
<td>0.151</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.782</td>
<td>-0.511</td>
<td>-0.149</td>
<td>-0.053</td>
<td>0.297</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>6</td>
<td>894</td>
<td>0.151</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.782</td>
<td>-0.511</td>
<td>-0.149</td>
<td>-0.053</td>
<td>0.297</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>7</td>
<td>894</td>
<td>0.151</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.782</td>
<td>-0.511</td>
<td>-0.149</td>
<td>-0.053</td>
<td>0.297</td>
<td>0.286</td>
<td>1.335</td>
</tr>
<tr>
<td>8</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>9</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>10</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>11</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>12</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>13</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>14</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>15</td>
<td>894</td>
<td>-0.056</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2.937</td>
<td>-0.511</td>
<td>-0.017</td>
<td>-0.343</td>
<td>0.297</td>
<td>0.794</td>
<td>1.335</td>
</tr>
<tr>
<td>16</td>
<td>894</td>
<td>0.013</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3.324</td>
<td>-0.235</td>
<td>0.744</td>
<td>-0.125</td>
<td>-0.642</td>
<td>-0.223</td>
<td>1.335</td>
</tr>
<tr>
<td>17</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>18</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>19</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>20</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>21</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>22</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>23</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>24</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>25</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>26</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>27</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>28</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>29</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>30</td>
<td>1010</td>
<td>1.388</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.162</td>
<td>-0.630</td>
<td>2.450</td>
<td>-0.779</td>
<td>2.175</td>
<td>2.827</td>
<td>1.335</td>
</tr>
<tr>
<td>010</td>
<td>1.388</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Problem.

- Predicting comorbidities at the earliest from time-series data is challenging.
- Each patient has a dynamic and unique profile.
- Comorbidities interact.
- Many unmeasured effects.
The solution, Personalising Medicine

Hidden Variable discovery approach
• Finding methods to assess the influences of these latent variables,
• Discover the dependencies between the latent variable and the observed variables.
• Discover Diabetic trigger and eliminate diabetes forever!
• Determining the precise position of the latent variable

• Our key contribution is the combination of the IC* algorithm to identify latent variables within Dynamic Bayesian Networks.
An example

Our hypothesis is:
“If Glycated Hemoglobin (HbA1c) is less than about (7%), then retinopathy may never develop, or develop very slowly.”

http://www.diabeticretinopathy.org.uk/prevention/hba1c_and_retinopathy.htm
Dynamic Bayesian Networks

- Ideal for clinical data:
  - Flexibility in continuous and discrete variable;
  - Handling uncertainty through the modelling of probability distributions;
  - Enables prediction through inference;
  - No limit for minimum sample size;
  - Transparent (querying the model, graphical structure)
  - It can naturally facilitate latent variables …
IC (Inductive Causation) algorithm

- It Applies conditional independence analyses to infer causal structures;
- IC* algorithm (an extension of IC) learns a partially oriented Directed Acyclic graph (pattern) with latent variables.

![Diagram showing causal relationships between variables a, b, and c.]

Whenever a then b but not vice versa
Possibly a => b

See Pearl, "Causality: Models, Reasoning, and Inference", 2000, p52 for more details.
Dynamic links

- Learning the temporal links of our DBNs, Using REVerse Engineering ALgorithm (REVEAL) (Liang1998)

- We assumed hidden variable status at time $t$ depends on the corresponding hidden variable at a previous time $(t-1)$. 
Pair sampling and Stepwise approach procedure

- Apply IC* algorithm on the balanced data.
- Provide probabilities of states by applying inference rules on all discovered hidden variables.
- Treat the discovered hidden variables as an observed variables.
- Re-apply the IC* and repeat all Steps until no new hidden variables are discovered.
- Having discovered the hidden variables, we build a predictive DBN model.
- Parameter estimation using the expectation-maximization (EM) algorithm.
Static structure with no hidden variable
Static structure with the first hidden variable
Step 2

Static structure with the second hidden variable
Static structure with the third hidden variable

Step 3
Confusion Matrix Results

The step-wise approach with a generally improving classification accuracy of diagnosing targeted complications as a number of hidden variables are added (the blue line represents retinopathy and the red line represents liver disease).

\[
G - \text{Mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}}
\]

\[
F - \text{Measure} = 2 \times \text{sensitivity} \times \text{precision} \times \text{sensitivity} + \text{precision}
\]
Hidden variable fluctuation

Predicted Latent Variable Pattern VS T2DM Complication and Features
Inference: Query Rules

QRS =

P (complication='being at risk of retinopathy' | Evidence)

Evidence =

{risk of having retinopathy is reduced = 'high level of hidden variable 3, lower level of hidden variable 2, and very low level of hidden variable 1'}

The variations in the latent variable are affected by various comorbidities.
Conclusion and Future Works

- Effectively integrates Bayesian methods with latent variables by adapting the prior probability of the event occurrence for future time points;
- The proposed method is more accurate than using one of hidden variable step or no hidden variables at all;
- Avoiding overfitting in the structure learning, using a stronger stopping rule in the step-wise approach;
- Exploiting mutual information metrics (Ebert, 2007) to filter some of the hidden variable relationships;
- Discovering interesting dependencies between the latent variable and the observed variables;
- Interpreting the impact of hidden (latent) variables in finding temporal phenotypes in the presence of unmeasured diabetic disorders;
- Concentrating on the continuous investigation of features;
- Exploring Deep Learning methods;
References


Thank you for listening!

Any Question?

“Type 2 diabetes is not going to kill me. I just have to eat right, exercise, lose weight, watch what I eat, and I will be fine for the rest of my life.”