Bayesian Methods for Diagnosing Physiological Conditions of Human Subjects from Multivariate Time Series Biosensor Data

Mehmet Kayaalp

Lister Hill National Center for Biomedical Communications
National Library of Medicine
National Institutes of Health
Department of Health and Human Services

Problem & Data

- **Diagnosis**: Given all (current and past) measures of a subject, determine subject’s current condition

- **Stochastic Process Modeling**

- **Multivariate Time Series Data**:  
  - Discrete Time
  - Continuous Covariate Data (Signal)
  - Discrete Response (Categories)
Outline

- Bayesian networks
  - Simple Bayesian networks
- Dynamic Bayesian networks
  - Stationary processes
  - First-order Markov processes
  - Dynamic simple Bayesian models
  - Modified dynamic simple Bayesian models
- Learning parameters from continuous data
Bayesian Networks

\[ BN = (S, \theta) \]
\[ S = (X, A) \]

- Components of Bayesian networks (\( BN \))
  - Structure \( S \) represented on directed acyclic graphs (DAG), and
  - Parameters \( \theta \) represented in conditional probability distributions.

- \( S \) comprises
  - a set of random variables \( (X = X_1, X_2, X_3) \),
  - a set of relationships \( A = \{A_{1,2} = (X_1, X_2), A_{1,3} = (X_1, X_3)\} \)

\[ \theta = \{P(X_1), P(X_2 \mid X_1), P(X_3 \mid X_1)\} \]

\[ P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \]
A stochastic process is a sequence of random variables indexed by time.

A dynamic Bayesian network (DBN) is a graphical model representing stochastic processes.

\[ DBN = (S, \theta) \]

\[ S = (X(t), A) \]
A Markov process is a stochastic process, whose future states depend only on the current state.

\[ P(X(t_n) | X(t_{n-1})) = P(X(t_n) | X(t_{n-1}), \ldots, X(t_1)) \]

**Examples**

- **Markov**: Overall health condition
- **Non-Markov**: Serum-Glucose Level
Stationarity

- A stationary process is a stochastic process, in which the statistical relationships between variables do not change over time.

\[ P(X(t_1), X(t_2), \ldots, X(t_n)) = P(X(t_{1+d}), X(t_{2+d}), \ldots, X(t_{n+d})) \]

\[ d = \pm 1, \pm 2, \ldots \]

- Examples
  - Stationary: Heart rate at resting state
  - Nonstationary: Stock market, atmospheric dynamics, certain arrhythmias, atrial fibrillation
First-order Markov assumption limits dependencies between two consecutive states.

- Stationarity fixes the structure and parameters over the entire timeline.
- Both assumptions yield DBNs with two time slices.
- Variable values on the first time-slice (shaded nodes) are given.
Dynamic Simple Bayes\(^{(1)}\) (DSB)

- Stationary process models
- First-order Markov process models
- Arcs are in reverse direction of time flow
- Contemporaneous variables are conditionally independent
- Easy to build
- Fast in inference

Forecasting vs. Diagnosis

- **DSBs are designed for forecasting purposes**
  - Forecasting: Given data, predict the future conditions
- **Workshop problems are diagnosis tasks**
  - Diagnosis: Given data, identify the current (past) condition
- **Problem: Identify current physiological condition (PC) of a test subject**
- **Test data comprise current PCs**
Atemporal vs. Temporal

- Contemporaneous information is almost always more informative than historical information
- Is historical (temporal) information necessary?
- Simple Bayesian (SB) network
  - a robust atemporal classifier; i.e.,
  - no temporal differences between SB variables
Modified DSB (mDSB)

- DSB augmented with SB \( \rightarrow \) mDSB
  - utilizes both contemporaneous and historical information
- A diagnostic tool
- As better than SB as valuable historical information
Modified DSB (mDSB)

- DSB augmented with SB ? mDSB
  - utilizes both contemporaneous and historical information
- A diagnostic tool
- As better than SB as valuable historical information
Inference

\[ P(Y, x_1, \ldots, x_n) = P(Y) \prod_{i=1}^{n} P(x_i | Y) \]

\[ P(Y|x_1, \ldots, x_n) \propto P(Y) \prod_{i=1}^{n} P(x_i | Y) \]

\[ P(Y|x_1, \ldots, x_n) \propto P(Y) \prod_{i=1}^{n} \frac{P(Y|x_i) P(x_i)}{P(Y)} \]
Parameter Learning

\[ P(Y = y_k \mid X = x) = ? \]

- **How can we learn parameters from continuous data with high variance?**
  - **Smoothing?**
    - **Moving Averaging?**
    - **Windowing?**
Parameter Learning

\[
N_{y_k|x} = \sum_{x_i \in X} \frac{n(y_k, x_i)}{1 + e^{\frac{a}{\text{scale}(X)}|x_i - x| - c}}
\]

\[
n(y_k, x_i): \text{frequency count}
\]

\[
\text{scale}(X) = \max(X) - \min(X)
\]

\[
P(y_k | x) = \frac{\alpha_k + N_{y_k|x}}{\alpha + \sum_{y \in Y} N_{y|x}}
\]

\[
\alpha_k: \text{priors, where } \alpha = \sum_k \alpha_k
\]

- Learned by using all \((y_k, x)\) data points in the training dataset
- Data points closer to \(x\) have more impact on \(P(y_k | x)\) than others
- Based on a sigmoid decay function
  - Weighted by \(d(x, x_i)\)
Conclusions

- Two methods were introduced
  - Modified dynamic simple Bayesian (mDSB) modeling for diagnosis
  - Online parameter learning from continuous data through sigmoid decay functions (sdf)
- Submitted results were obtained through simple Bayesian network (SB) and mDSB topologies with parameters learned through sdf