Machine Learning for Healthcare
HST.956, 6.S897

Lecture 6: Physiological time-series

David Sontag
Outline of today’s lecture

1. Recap of risk stratification
2. Physiological time-series
   – Monitoring babies in neonatal ICUs
   – Detecting atrial fibrillation
Survival modeling with right-censored data

Event occurrence e.g., death, divorce, college graduation


© ACM. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
Notation and formalization

• $f(t) = \text{be the probability of death at time } t$
• Survival function: $S(t) = P(T > t) = \int_{t}^{\infty} f(x) dx$

Fig. 2: Relationship among different entities $f(t), F(t)$ and $S(t)$.

[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]
Maximum likelihood estimation

- Two kinds of observations: censored and uncensored

Uncensored likelihood

\[ p_\theta(T = t | x) = f(t) \]

Censored likelihood

\[ p_{\theta}^{\text{censored}}(t | x) = p_\theta(T > t | x) = S(t) \]

- Putting the two together, we get:

\[ \sum_{i=1}^{n} b_i \log p_{\theta}^{\text{censored}}(t | x) + (1 - b_i) \log p_\theta(t | x) \]

Optimize via gradient or stochastic gradient ascent!
Evaluation for survival modeling

• Concordance-index (also called C-statistic): look at model’s ability to predict relative survival times:

\[ \hat{c} = \frac{1}{\text{num}} \sum_{i: \delta_i = 0} \sum_{j: y_i < y_j} I[S(\hat{y}_j|X_j) > S(\hat{y}_i|X_i)] \]

• Illustration – blue lines denote pairwise comparisons:

Black = uncensored
Red = censored

• Equivalent to AUC for binary variables and no censoring

Final thoughts on survival modeling

• Could also evaluate:
  – Mean-squared error for uncensored individuals
  – Held-out (censored) likelihood
  – Derive binary classifier from learned model and check calibration

• Partial likelihood estimators (e.g. for cox-proportional hazards models) can be much more data efficient
Dealing with non-stationarity

• Baseline: Retrain the model with most recent data

• How to best use historical data?
  – Impute or transform historical data to look like current data (e.g., Ganin et al., JMLR ‘16)
  – Reweight historical data to look like current data (see e.g. Sugiyama and Kawanabe, ‘12)
  – Online algorithm that adapts quickly (see e.g. Shen et al. AI Stats ‘18)
Recap of risk stratification

• Classification vs. survival modeling (regression)
• Causal interpretation of predictive features
• Imputation of missing data
Outline of today’s lecture

1. Recap of risk stratification

2. Physiological time-series
   - Monitoring babies in neonatal ICUs
   - Detecting atrial fibrillation
Physiological time-series

Fig. 4. Probes used to collect vital signs data from an infant in intensive care. 1) Three-lead ECG, 2) arterial line (connected to blood pressure transducer), 3) pulse oximeter, 4) core temperature probe (underneath shoulder blades), 5) peripheral temperature probe, 6) transcutaneous probe.

© IEEE. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

(Quinn et al., TPAMI 2008)
Physiological time-series

• Typical use cases:
  1. Infer true physiological signal from noisy observations
  2. Risk stratification, e.g. predict clinical deterioration, or diagnosis

• Approach taken depends on:
  – Is labeled data available?
  – Do we have a good mechanistic/statistical model?
  – How much training data is there?
Two very different trajectories

(Queen et al., TPAMI 2008)
Problem: measurements confounded by interventions & measurement errors

**Blood pressure**
- Sys. BP (kPa)
- Dia. BP (mmHg)

**Oxygen uptake**
- TcPO$_2$ (kPa)
- TcPCO$_2$ (kPa)

**BS**
- TR

© IEEE. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [https://ocw.mit.edu/help/faq-fair-use/](https://ocw.mit.edu/help/faq-fair-use/)

(Quinn et al., TPAMI 2008)
Can we identify the artifactual processes?

• Once identified, can remove for use in downstream predictive tasks (must deal with missing data)
• Can help mitigate *alarm fatigue* by not alerting the clinicians when unnecessary
• More broadly, can we maintain beliefs about the true physiological values of a patient?
(Switching) linear dynamical systems

- Conditioned on $s_t$, linear Gaussian state-space models (Kalman filters):
  \[
  x_t \sim \mathcal{N} \left( A^{(s_t)} x_{t-1} + d^{(s_t)}, Q^{(s_t)} \right)
  \\
y_t \sim \mathcal{N} \left( C^{(s_t)} x_t, R^{(s_t)} \right)
  \]
(Switching) linear dynamical systems

• Full model:

\[
\begin{align*}
\mathbf{s}_{t-1} & \rightarrow x_{t-1} \rightarrow x_t \rightarrow x_{t+1} \\
\mathbf{s}_t & \rightarrow y_{t-1} \rightarrow y_t \rightarrow y_{t+1} \\
\mathbf{s}_{t+1} & \rightarrow \mathbf{y}_t \\
\end{align*}
\]
Learning SLDS models

- Assume some labeled training data \( \{s, y\} \)
- True state \( x \) assumed to never be observed
- Learn using expectation maximization
Parameterizing model

- Normal heart rate dynamics are well-modeled using an autoregressive process, e.g.

\[
\begin{align*}
x_t & \sim N \left( \sum_{k=1}^{p_1} x_k (x_{t-k}, b_{t-k}), \eta_1 \right) \\
b_t & \sim N \left( \sum_{k=1}^{p_2} b_{t-k}, \eta_2 \right)
\end{align*}
\]

Baseline process (smooth)  
Zero-mean, high frequency

© IEEE. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [https://ocw.mit.edu/help/faq-fair-use/](https://ocw.mit.edu/help/faq-fair-use/)  

(Quinn et al., TPAMI 2008)
Parameterizing model

• One can use domain knowledge to specify parts of the artifacts model
  – Probe dropouts modeled by removing dependence of observation $y_t$ on patient state $x_t$
  – Temperature probe disconnection: exponential decay to room temperature
Evaluation

- 3-fold cross validation, where for each fold train on 10 babies and test on 5
- 24-hours of data for each baby
- Normal dynamics refit for test babies using a 30-minute section near the start

(Quinn et al., TPAMI 2008)
Evaluation

GS = Gaussian-sum approximation (used for inference)

RBPF = Rao-Blackwellized particle filtering approximation (used for inference)

FHMM = Factorial HMM (simpler model which does not model normal physiological dynamics)

(Quinn et al., TPAMI 2008)
Inference of physiological state

Blood sample draw

Temperature probe disconnection

© IEEE. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/

(Quinn et al., TPAMI 2008)
Inferred switch settings

TD = core temperature probe disconnection
TR = recalibration

© IEEE. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
Outline of today’s lecture

1. Recap of risk stratification
2. Physiological time-series
   – Monitoring babies in neonatal ICUs
   – Detecting atrial fibrillation
Detecting atrial fibrillation

AliveCore ECG device

ECG = electrocardiogram
What type of heart rhythm?

- Normal rhythm
- AF rhythm
- Other rhythm
- Noisy recording
Traditional approach

2. Common structure of the QRS detectors.

Winning approach

• Training data in 2017 Physionet challenge: ~8500 ECGs
• Best algorithms use a combination of expert-derived features and machine learning

Not enough data for deep learning? Wrong architectures?

“However, the fact that a standard random forest with well chosen features performed as well as more complex approaches, indicates that perhaps a set of 8,528 training patterns was not enough to give the more complex approaches an advantage. With so many parameters and hyperparameters to tune, the search space can be enormous and significant overtraining was seen...”

[Clifford et al. AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge, Computing in Cardiology 2017]
Differences with previous work

- Sensor is a Zio patch – conceivably much less noisy:

- ~90K ECG records annotated (from ~50K patients)
- Identify 12 heart arrhythmias, sinus rhythm and noise for a total of 14 output classes
Deep convolutional network

- 1-D signal sampled at 200Hz, labeled at 1 sec intervals
- 34 layers
- Shortcut connections (ala residual networks) with max-pooling
- Subsampled every other layer (2^8 in total)

Summary

• We are nearly always in realm of “not enough data”
• Modeling and incorporating prior knowledge is critical to good performance
• Design principles
  – Model the distribution of physiological dynamics
  – Derive features using existing clinical knowledge
  – Start from the simplest possible model
  – Share statistical strength across tasks