Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier

Joseph Futoma[1], Sanjay Hariharan[1], Katherine Heller[1]

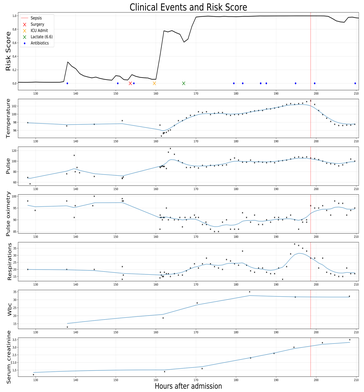
[1]Dept. of Statistical Science, Duke University

Background

- **Sepsis**: inflammatory response and complication arising from infection, with high morbidity and mortality.
- **Early intervention improves patient outcomes!**
  - Mortality from septic shock increases 7.6% each hour treatment is delayed after hypotension.
  - Mortality odds ratio 1.04 (95% CI: (1.03,1.06)) for each hour antibiotics delayed.
- Early, accurate identification of sepsis is challenging, as symptoms may be caused by many other conditions.
- Many data-driven "early warning scores" exist in medicine, e.g. NEWS.
  - Compare a small number (e.g. 7) physiological variables to normal ranges.
  - Sum score for each variable to get a composite score.
- These scores have low precision, high alarm fatigue.
  - Duke tried NEWS for sepsis, but 63.4% of alerts were cancelled.
- **Goal**: More flexible statistical model, use all available data to predict early onset of sepsis.

Data

- Electronic Health Records (EHRs): capture patient information from encounters with the health system.
- In our work we use the following structured variable types:
  1. 35 baseline covariates: Demographics, admission status (e.g. transfer, emergency), comorbidities.
  2. 8 Medication classes: administration times when patient was treated
  3. 34 physiological time series: 5 vitals, 29 laboratory test results
- 49,312 inpatient encounters over 18 months. Mean length of stay 121.7 hours (sd 108.1).
- Diagnosis codes are poor (no timestamp), no standardized definition, no biomarker.
- The "sepsis event" time that we will model is the first time that the following 3 events occur:
  1. At least 2/4 persistent abnormal vitals.
  2. Evidence of organ failure from abnormal lab result.
  3. Blood culture ordered for suspected infection.
- Final dataset has 21.4% of encounters with a sepsis event time.



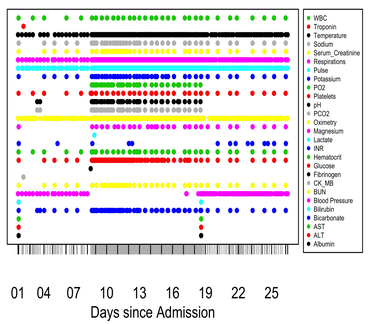


Figure 1: **Left**: model risk score over time for an example encounter where sepsis would have been detected very early. **Right**: visualization of irregular spacing of observation times for this encounter.

Proposed Model

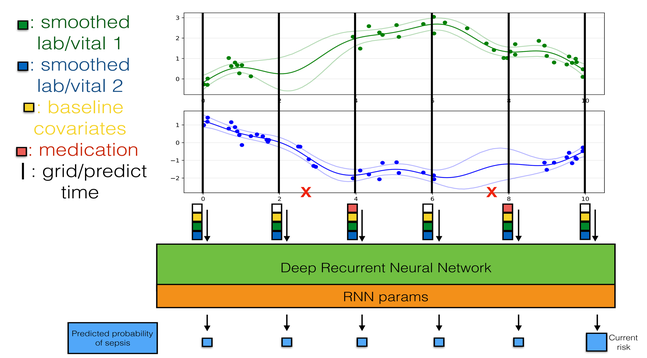


Figure 2: Multitask Gaussian Process imputes, smooths raw labs/vitals. Samples fed with baseline covariates, medications into RNN to predict sepsis.

- Multivariate time series classification: update probability an encounter is septic with streaming data.
- **Multitask Gaussian Process (MGP) for Clinical Time Series**:
  - $f_{im}(t)$: latent function, true values of variable m for patient encounter i at time t.
  - MGP model: GP priors over f_{im}, with shared correlation function k^t over time:

$$\text{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t') \quad (1)$$

$$y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m). \quad (2)$$

  - $y_{im}(t)$: observed value. $T_i \times M$ matrix $\mathbf{Y}_i$: fully observed series of M measurements at T_i times:

$$\text{vec}(\mathbf{Y}_i) \equiv \mathbf{y}_i \sim \mathcal{N}(0, \Sigma_i) \quad (3)$$

$$\Sigma_i = K^M \otimes K^{T_i} + D \otimes I. \quad (4)$$

    - K^M: $M \times M$ covariance matrix among lab/vital variables.
    - K^{T_i}: $T_i \times T_i$ correlation matrix for observation times $\mathbf{t}_i$, from correlation function k^t, parameters η shared.
    - OU kernel: $k^t(t,t') = e^{-|t-t'|/l}$, with a single length-scale parameter $\eta = l$.
    - D: diagonal matrix of noise variances $\{\sigma^2_m\}_{m=1}^M$.
  - Model handles irregular spacing, missing values in raw data, and outputs a more uniform representation with **uncertainty**.
  - $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iX_i})$: evenly spaced points in time (e.g. every hour).
  - $X_i \times M$ matrix $\mathbf{Z}_i$ of latent time series values at the grid times $\mathbf{x}_i$. MGP provides normal posterior for $\mathbf{z}_i = \text{vec}(\mathbf{Z}_i)$:

$$\mu_{z_i} = (K^M \otimes K^{X_i T_i}) \Sigma_i^{-1} \mathbf{y}_i \quad (5)$$

$$\Sigma_{z_i} = (K^M \otimes K^{X_i}) - (K^M \otimes K^{X_i T_i}) \Sigma_i^{-1} (K^M \otimes K^{T_i X_i}) \quad (6)$$

    - $K^{X_i T_i}$, K^{X_i}: correlation matrices between $\mathbf{x}_i$, $\mathbf{t}_i$ and between $\mathbf{x}_i$ with itself, from correlation function k^t.
    - MGP parameters to learn: $\theta = (K^M, \{\sigma^2_m\}_{m=1}^M, \eta)$.
- **Classification with an RNN**:
  - Latent values $\mathbf{z}_i$ at shared times $\mathbf{x}_i$ are now standardized inputs to the classifier.
  - Classifier: deep recurrent neural network, Long-Short Term Memory (LSTM) architecture, as $\mathbf{z}_i$ and $\mathbf{x}_i$ are variable size.
  - At x_{ij}, $\mathbf{d}_{ij} = [\mathbf{z}_{ij}^\top, \mathbf{b}_i^\top, \mathbf{p}_{ij}^\top]^\top$ fed into RNN: latent function values $\mathbf{z}_{ij}$, baseline covariates $\mathbf{b}_i$, medication indicators $\mathbf{p}_{ij}$.
  - **Problem**: $\mathbf{z}_i$ never directly observed, since $\mathbf{z}_i$ is random variable! Given $\mathbf{z}_i$, learning is standard.
  - Optimize an expected loss function, with respect to posterior on $\mathbf{z}_i$. Learning problem:

$$w^*, \theta^* = \text{argmin}_{w,\theta} \sum_{i=1}^N \mathbb{E}_{z_i \sim N(\mu_{z_i}, \Sigma_{z_i}, \theta)}[l(f(\mathbf{D}_i; w), o_i)]. \quad (7)$$

    - $f(\mathbf{D}_i; w)$: RNN classifier function, parameterized by w, maps input matrix $\mathbf{D}_i$ to predicted probability.
    - $l(f(\mathbf{D}_i; w), o_i)$: standard RNN loss function (e.g. cross entropy), comparing prediction to binary label o_i.
  - Given w^*, θ^*, to predict for a new patient, $\mathbb{E}_{z_j \sim N(\mu_{z_j}, \Sigma_{z_j}, \theta^*)}[f(\mathbf{D}_j; w^*)]$ is a risk score to update continuously.
  - Approach is "**uncertainty-aware**": uncertainty in MGP posterior for z_i propagated through to the loss.
- **End-to-End Learning**:
  - Expected loss $\mathbb{E}_{z \sim N(\mu_z, \Sigma_z, \theta)}[l(f(\mathbf{D}; w), o)]$ intractable, approximate with Monte Carlo samples:

$$\mathbb{E}_{z \sim N(\mu_z, \Sigma_z, \theta)}[l(f(\mathbf{D}; w), o)] \approx \frac{1}{S} \sum_{s=1}^S l(f(\mathbf{D}_s; w), o), \quad (8)$$

$$\mathbf{D}_s = [\mathbf{Z}_s^\top, \mathbf{B}^\top, \mathbf{P}^\top]^\top, \;\; \text{vec}(\mathbf{Z}_s) \equiv \mathbf{z}_s \sim N(\mu_z, \Sigma_z; \theta) \quad (9)$$

  - To compute gradients, use reparameterization trick: $\mathbf{z} = \mu_z + R\xi$, with $\xi \sim N(0, I)$ and $\Sigma_z = RR^\top$.
  - R: matrix square root. Lanczos method to approximate $\Sigma_z^{1/2}\xi$, yielding cheap draws for $\mathbf{z}$.
  - Conjugate gradient: speed up computation of μ_z, and matrix-vector products with Σ_z.
  - All operations in conjugate gradient, Lanczos are differentiable, so can use backpropagation.

Results

- Throw away data after sepsis; for non-sepsis cases use all data.
  - 80%/10%/10% Training/Validation/Testing split of data.
- Evaluate with AU-ROC, False Alarms per True Alarm; vary how far in advance prediction made.
- Baseline methods to compare with our method, MGP-RNN:
  - MGP-RNN-mean: Replaces $\mathbf{z}$ with its mean, μ_z. Faster but discards uncertainty.
  - GP-RNN-indep: Replace MGP with univariate GP for each variable, each with own kernels.
  - GP-RNN-shared: Replace MGP with univariate GP for each variable, with shared kernels.
  - Raw RNN: RNN classifier on raw data. To impute, carry last observation forward.
  - PLR: L_1 penalized logistic regression, same imputation as Raw RNN.
  - NEWS: clinical early warning scores previously used at Duke.

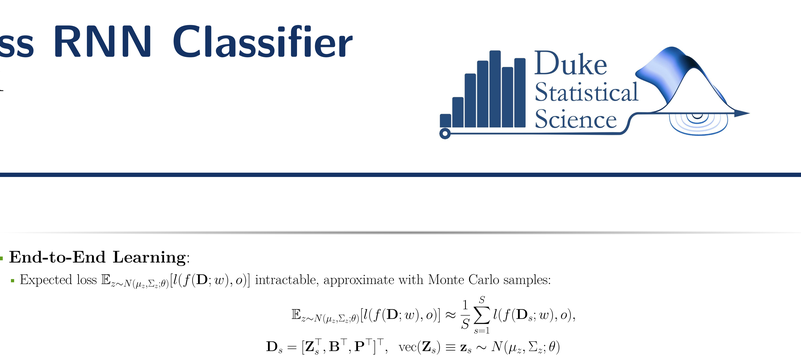


Figure 3: Results, as function of hours in advance predictions are made. **Left**: AU-ROC. **Right**: False Alarms per True Alarm, at 85% Sensitivity.

- 4 variations on our method in solid; substantially outperform other baselines at close to sepsis.
- MGP-RNN has modest improvements over the mean version, univariate GP baselines.

Conclusion

- Novel approach for early detection of sepsis with excellent performance.
- New framework for classification of noisy, irregular clinical time series.
- Strongly outperforms NEWS, clinical baseline used at Duke; much higher precision.
- Extensions:
  - Apply to other clinical conditions (e.g. code blue prediction).
  - Learn treatment-response curves from medications.
  - More sophisticated covariance structure in MGP, e.g. Linear Model of Coregionalization.
  - More advanced classifiers, e.g. variational RNNs, or attention mechanism for interpretability.
- **Code is available! https://github.com/jfutoma/MGP-RNN**