An Improved Multi-Output Gaussian Process RNN with Real-Time Validation for Early Sepsis Detection

Joseph Futoma[1], Sanjay Hariharan[1], Katherine Heller[1], Mark Sendak[2], Nathan Brajer[2], Meredith Clement[3], Armando Bedoya[3], Cara O'Brien[3]

Duke University: [1]Dept. of Statistical Science [2] Institute for Health Innovation [3] Dept. of Medicine

Background

- **Sepsis**: complication from infection, high morbidity and mortality.
- **Early intervention improves patient outcomes!**
  - Mortality odds ratio 1.04 (95% CI: (1.03,1.06)) for each hour antibiotics delayed.
- Early, accurate identification of sepsis is hard!
- Many data-driven "early warning scores" exist in medicine, e.g. NEWS.
  - Compare a small number (e.g. 7) physiological variables to normal ranges.
  - Low precision, high alarm fatigue (63.4% of NEWS alerts cancelled at Duke for sepsis).
- Our prior work combines Multitask Gaussian Processes, RNNs to detect sepsis.
- **Goal**: Improve existing MGP-RNN, come up with better validation scheme.

Data

- Structured variable types we extracted from the EHR:
  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.
- 51,697 inpatient encounters over 18 months. Mean length of stay 121.7 hours.
- **Problem**: how to identify when sepsis occurred from real EHR data?
- Diagnosis codes are poor (no timestamp), no standardized definition, no biomarker.
- "Sepsis event": first time 3 events occur (analogous to "Severe Sepsis"):
  1. At least 2/4 persistent abnormal vitals (determined by SIRS score).
  2. Evidence of organ failure from abnormal lab results.
  3. Blood culture ordered for suspected infection.
- Final dataset has 21.4% of encounters with a sepsis event time.

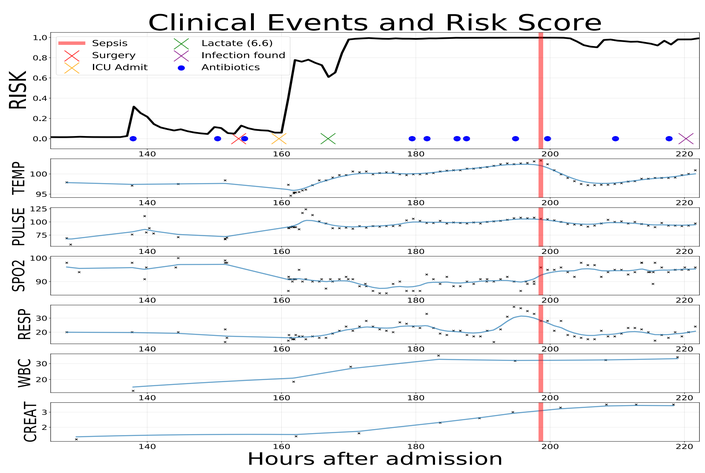

Figure 1: Model risk for an example encounter where sepsis would have been detected very early.

Previous MGP-RNN

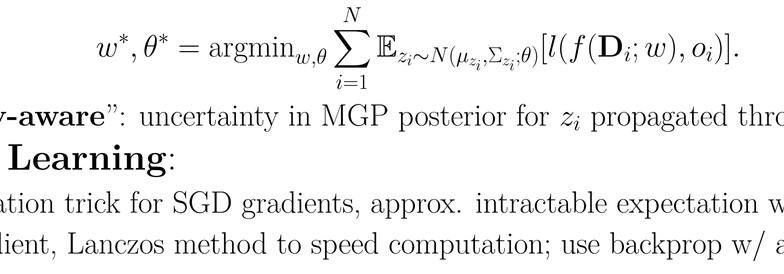

Figure 2: Multi-output GP imputes, smooths raw data. Samples fed w/ static covs., meds into RNN.

- Time series classification: update sepsis probability with streaming data.
- **Multi-output Gaussian Process (MGP) for Clinical Time Series**:
  - $f_{im}(t)$: latent value, lab m, patient i, time t; $y_{im}(t)$: observed value.
  - Multitask GP: GP priors over f_{im}, shared correlation function k^t over time:

$$\text{cov}(f_{im}(t), f_{im'}(t')) = K^M_{mm'} k^t(t,t'), \quad y_{im}(t) \sim \mathcal{N}(f_{im}(t), \sigma^2_m).$$

  - Handle irregular spacing, missing values, outputs uniform representation with **uncertainty**.
  - MGP provides posterior distribution for latent function values z_i on evenly spaced grid (Fig. 2).
- **Classification with an RNN**:
  - Inputs to RNN classifier at grid times: latent function values, static covs., med indicators.
  - Since latent function values z_i are random, optimize an expected loss function w.r.t. z_i:

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

  - "**Uncertainty-aware**": uncertainty in MGP posterior for z_i propagated through to the loss.
- **End-to-end Learning**:
  - Reparameterization trick for SGD gradients, approx. intractable expectation w/ MC samples.
  - Conjugate gradient, Lanczos method to speed computation; use backprop w/ auto diff.

Extenstions to the MGP-RNN

- **Increasing Flexibility of MGP**:
  1. Relax zero-mean prior, let mean function depend on past meds given:

$$\mu_m(t) = \sum_{p=1}^{P} \sum_{t_p<t} f_{pm}(t - t_p), \quad f_{pm}(t) = \sum_{l=1}^{L} \alpha_{lpm} e^{-\beta_{lpm} t}$$

  2. More flexible kernel: sum of Q separable kernels ($Q = 1$ is Multitask GP).

$$\text{cov}(f_{im}(t), f_{im'}(t')) = \sum_{q=1}^{Q} K^M_q(m,m') k^t_q(t,t')$$

- **Improvements to the RNN**:
  1. Target Replication: RNN loss depends on multiple outputs (2 hrs before, 6 after sepsis).
     - Help alleviate issues with label ambiguity by labelling multiple time points during training.
  2. Missingness Indicators: model missing data pattern, input to RNN indicators for labs sampled.
     - RNN now learns complex interactions between lab/vital values, lab/vital sampling times, meds, baseline covariates.

Results

- Case-Control Matching: create more realistic, harder prediction problem.
  - Previously: compare sepsis cases prior to sepsis w/ controls prior to discharge. Too easy!
  - Now: match each sepsis patient to 4 non-sepsis patients w/ similar LOS, static covs.
  - For matched controls, "prediction time" is same % through LOS as sepsis time for matched case.
- Evaluation: see how metrics vary vs hours prior to sepsis / "prediction time".
- Baselines: RNN w/ LOCF imp., Random Forest, Pen. Log. Reg., Clinical EWSs.

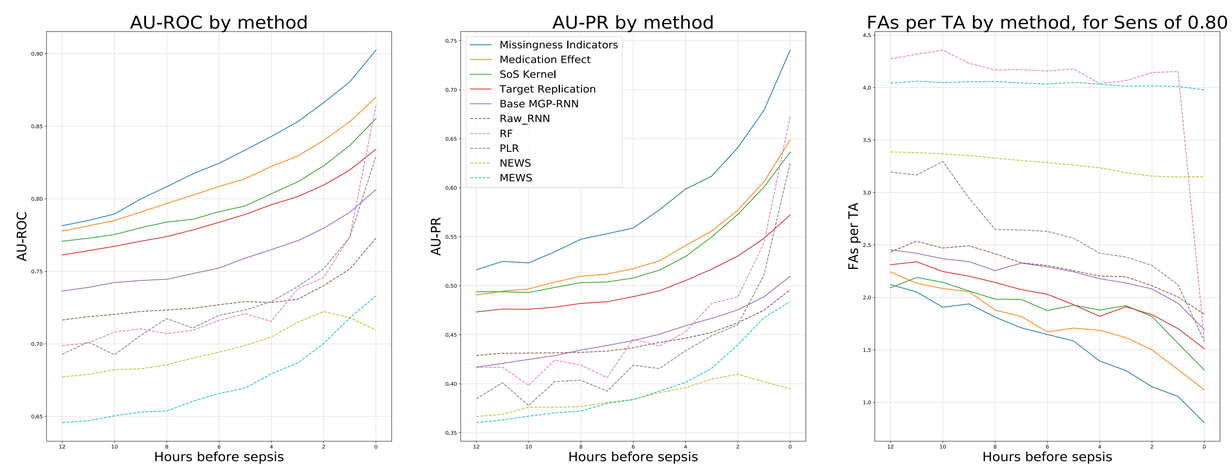

Figure 3: "Lookback" Results on Matched Data, as function of hours before sepsis / "prediction time". **Left**: AU-ROC. **Middle**: AU-PR **Right**: False Alarms/ True Alarm, 80% Sens.

- Real-Time Validation: want evaluation that more closely mirrors actual use case.
  - "Lookback" results require alignment at end of time series; in practice, not known in advance.
  - Across all $[0, 1]$ thresholds, get time of first alert per patient; create confusion matrices for metrics.

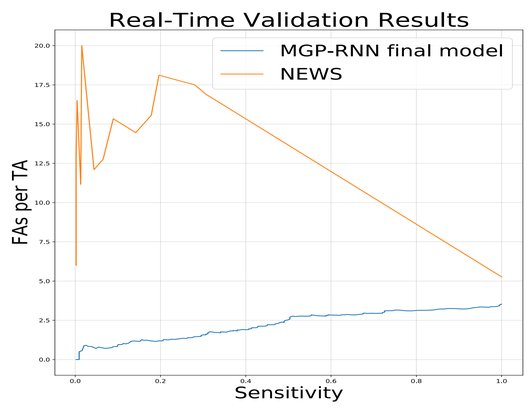

Figure 4: "Real-Time" Results: False Alarms / True Alarm as function of Sens.

Conclusion

- Improved framework for classification of noisy, irregular clinical time series.
- Developed a more practical, general-purpose real-time validation scheme.
  - Our model strongly outperforms NEWS previously used at Duke; much higher precision.
- Our approach will be validated in a **Randomized Controlled Trial** this fall!
- **Code is available! https://github.com/jfutoma/MGP-RNN**

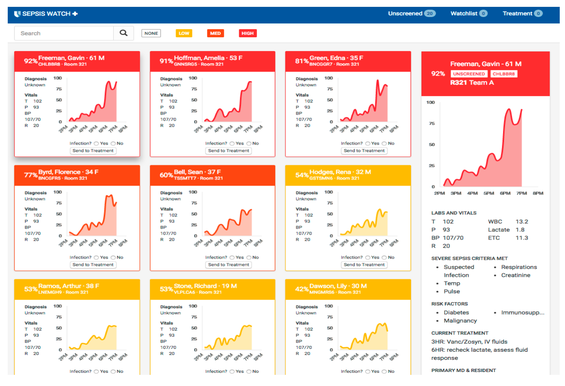

Figure 5: App developed to display model predictions and other data to be used by Sepsis RRT.