Personalized infection probability.

Healthcare associated infections (HAI) are estimated to cost US hospitals $9.8B per year. Timely assessing risk of infection would allow clinicians and infection specialists to prevent infections while saving costs. In this study we are interested in estimating conditional probabilities of infection given their sequence (history) of procedures, individual characteristics and characteristics of the unit they are staying in.

\[ P(\text{infection} = 1 | \text{Characteristics of patient, History of procedures of patient, Unit characteristics}) \]

Pre-processing data as sequences.

The problem is challenging as the estimation of conditional probabilities is limited by a dataset, and a new patient with unseen joint characteristics or procedures would not have an estimated probability of infection. Moreover, as the order of procedures matter, the simulation must store information about the previous history of procedures. To achieve this, we encode the data into a sequence of integers.

Building a Markovian simulation scheme from non-Markovian sequences.

With input data as non-Markovian sequences of integers \( x(t) \), we formulate a parametrized estimation of a joint Markovian process \((x(t), h(t))\). Where \( h(t) \) represents an enlargement of the state space of \( x(t) \), such that the joint process is Markovian:

\[
(x_{t+1}, h_{t+1}(\theta)) = f((x_t, h_t(\theta)), U)
\]

The parameters \( \theta \) governing the memory process \( h(t) \) are optimized in the training procedure and are modeled with a Recurrent Neural Network (RNN).

Parametrizing the memory process \( h(t) \) as a Recurrent Neural Network (RNN).

The complete pipeline for prediction is:

\[
\begin{align*}
\text{Input Integer} & \quad \rightarrow \quad \text{Embedding Layer} \quad \rightarrow \quad \text{GRU} (E, h_t) \quad \rightarrow \quad \text{Dropout Layer} \quad \rightarrow \quad \text{Vector of Probabilities for } X_{t+1}
\end{align*}
\]

For the RNN architecture we chose a Gated Recurrent Unit (GRU) network below. Alternatively, GRU can be replaced with a plain RNN or a Long-short Term Memory (LSTM) architecture.

Once trained, with fixed parameters \( \theta \), we can simulate paths of procedures performed on a patient until discharge.

Conclusions

Our model can predict and generate sequences of patient events conditional on any given set of previous events. Moreover, the model can be used to predict other metrics of interest such as probability of death and expected time to discharge. The rates of infection implied by the model match the real ones (left), and the model can be used to study how certain procedures affect the risk of infection over time (right).

Funding

Agency for Healthcare Research and Quality, AHRQ-R01-HS024915-01 (PI: Elaine Larson)
Columbia University Medical Center, School of Nursing

Collaborators

Elaine Larson, Philip Zachariah, Bevin Cohen, Jianfang Liu, Jingjing Shang (CUMC); Weichong Ni (IEOR).