Predicting Sick Patient Volume In a Pediatric Outpatient Setting using Time Series Analysis
Grace Guan, Barbara E. Engelhardt
Department of Computer Science, Princeton University

The Problem: Medical Wait Times
2017 Study of Outpatient Visits in USA
• 20% visits >20 min wait; 10% visits >30 min wait
2015 Study of London Hospitals
• 3.6% of deaths avoidable if treatment were faster

Background
We work with a pediatric group (PG) in the tri-state area.
• They serve 30,000 yearly patients and have 2 types of appointments:
  - Well Visits: Booked around 3 months in advance (check-ups)
  - Sick Visits: (think urgent care) Booked no more than 24 hours in advance (walk-ins)
• There is a constant amount of well visits scheduled each day; remainder of doctors’ time intended for sick visits.
• The PG has low reviews on Yelp because of lack of well visit availability and long wait times.

Five Models
• Constant Prediction (currently employed)
• Linear Regression (baseline)
• Time Series SARIMA (TS)
• Non-linear ReLU neural network (NN)
• Recurrent LSTM neural network (RNN)

Three Time Scales for Prediction
• Dynamic = Train on 2010-2016 data; validate or withhold 2017 data; predict held out 2018 data
• One Month Ahead (OMA) and One Step Ahead (OSA) = Train on 2010-2016 data; walk-forward validation by month or day respectively through 2017 and 2018

Features include calendar, weather, and lag variables
Labels include sick visits per full time pediatrician (per FTP), sick visits with a respiratory diagnosis per FTP, sick visits with a non-respiratory diagnosis per FTP

Data
• Data are evidently not constant over time. Graphing all data illustrates seasonal cycles as well as difference between non-holiday weekdays (black), Saturdays (green), Sundays (blue), and holidays (red).

Graphing data over 7.5 weeks reveals weekly cycles which follow a “u” shape pattern. Labor day, in red, is the first Monday in September.

Dynamic Model Performance
• We test our dynamic model predictions on held out 2018 data. The RNN performs best on this prediction task, with MSE 1.085.
• All models under-predict visits in February, and over-predict throughout March and April.
• This suggests more recent and detailed information should be captured in our models.

Conclusion & Future Work
• RNN models best capture seasonality of the data and can be used to perform outbreak detection by identifying error outliers.
• There are improvements in prediction when modeling sick patients as a mixture of disease types and when more recent data is used, through lag features and walk-forward validation.
• Future work should investigate whether other clinical settings that do not have major seasonal components (oncology, orthopedics) also find RNN models useful for demand forecasting.

Acknowledgments
We work with a pediatric group (PG) in the tri-state area.
• They serve 30,000 yearly patients and have 2 types of appointments:
  - There is a constant amount of well visits scheduled each day; remainder of doctors’ time intended for sick visits.
  - The PG has low reviews on Yelp because of lack of well visit availability and long wait times.

References