Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

ANALYTICAL AND SIMULATION-DRIVEN MACHINE LEARNING METHODS FOR GENERATING REAL-TIME OUTPATIENT LENGTH-OF-STAY PREDICTIONS

Najiya Fatma
Varun Ramamohan

Department of Mechanical Engineering
Indian Institute of Technology Delhi
Hauz Khas
New Delhi, 110016, INDIA

ABSTRACT

In this work, we consider real-time prediction of lengths-of-stay (LOS) for outpatients at a primary healthcare (PHC) facility via two methods: an analytical queuing-theoretic predictor, and simulation-driven machine learning (SimML) predictors. These LOS predictions are made at the point in time at which the patient is expected to arrive at the facility (i.e., at time $t + \delta$), using the system state of the PHC recorded at current time t. We develop a discrete-event simulation (DES) of operational flows of outpatients, inpatients and childbirth patients treated at the PHC. Both the analytical and SimML predictors use real-time system state information such as the number of patients waiting and elapsed service times of patients undergoing service. SimML predictors are trained using system state data generated by the DES for each outpatient. LOS predictions can inform patient decisions regarding which facility to visit and equitably manage resource utilization across a network of similar facilities.

1 INTRODUCTION

Primary health centers (PHCs) are the first point of contact for an Indian patient with the public healthcare delivery system, and they provide outpatient general medicine care, and limited inpatient and childbirth care. There is evidence of overcrowding and significant delays in accessing healthcare at certain PHCs owing to high patient load and patient preferences for a subset of facilities in region, leading to inequitable utilization of the network of such facilities in a given region. Consequently, predicting expected LOS estimates at healthcare facilities can help patients make an informed decision regarding which facility to visit (e.g., the one with the smallest LOS) and can help reduce disparities in utilization of these facilities.

We develop two approaches towards generating LOS estimates for outpatients seeking to visit PHCs on a real-time basis: (a) an analytical predictor grounded in queuing theory that uses system state data, and (b) a simulation-driven machine learning (ML) approach that involves using system state data for training ML methods for predicting outpatient LOS. We refer to the second approach as SimML predictors. In both approaches, outpatient LOS is predicted using system state information recorded at current time t – such as the number of patients waiting and the elapsed service time of patients in service at each facility subunit – for outpatients expected to arrive at the facility at time $t + \delta$. We record these operational information using a validated simulation model of the queuing system represented by the outpatients. Previous studies for LOS prediction predominantly applied data mining techniques to extract information regarding patient characteristics and admission and discharge dates of patients in emergency departments, and then trained ML methods on the extracted data to generate LOS estimates. We develop the second approach for a case where such historical data - system state or otherwise - is not available for training ML methods. Estimating LOSs using SimML technique will be significant for complex queuing systems encountered in healthcare facilities where developing a closed form expression for the analytical LOS predictor is difficult. We compare the performance of the analytical and SimML LOS predictors via a regression performance measure called as the Mean Absolute Percentage Error (MAPE).

2 LOS PREDICTION METHODOLOGY

We developed a DES model of patient care operations (Figure 1) using the *salabim* Python package for DES. Each patient type has an exponential interarrival time and a generally distributed service time with nonpreemptive priority for childbirth patients and inpatients over outpatients for receiving service from the doctor. We refer readers to Fatma et al. (2020) for a detailed description of PHC simulation development. The analytical predictor in Figure 2 generates real-time LOS predictions in steps. First, real-time delay predictions at each queuing subsystem within the PHC are generated. Service times at each subsystem are then added to expected delay estimates to determine real-time LOS estimates at each subsystem. These LOS estimates are summed over all subsystems to generate the total expected LOS estimates.

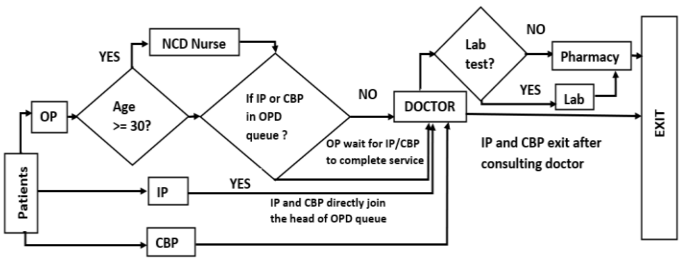

Figure 1: Patient flow at PHC. OP: Outpatients, IP: Inpatients, CBP: Childbirth patients.

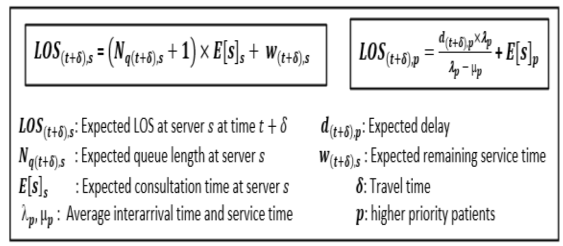

Figure 2: LOS predictor.

Now, for the SimML approach, for each outpatient, we generated system-state information from each PHC subsystem (a total of 16 features) via the DES. We generated data for 10,000 outpatients, and partitioned into training/testing sets in a 4:1 ratio. For each SimML predictor, we identified optimal hyperparameters using the *hyperopt* Python package. We also trained a feedforward neural network with three hidden layers with 64, 96, and 160 neurons, respectively, a learning rate of 0.001, and the $tanh$ activation function.

3 RESULTS

We ran the simulation for a period for 365 days of PHC operations (warm-up = 180 days) for generating the training data for the SimML predictors and for validating the analytical predictor. We quantify regression accuracy using MAPE score. We trained a variety of ML methods – as regressors (as opposed to classifiers) - on the dataset, ranging from feedforward neural networks, random forest, XGB, KNN regressor. We report the mean and standard deviation of the MAPE scores of each regressor, including the analytical predictor, in Table 1.

Table 1: MAPE scores.

Metric	**Analytical predictor**	**Random forest**	**Gradient boosting**	**KNN**	**XGB**	**ANN**
MAPE	10.40 (0.41)	14.09 (0.04)	7.33 (0.01)	4.73 (0.04)	1.88 (0.01)	8.78 (0.22)

4 CONCLUSIONS

In this work, we provide a methodology for predicting LOS estimates of outpatients using both analytical and simulation-driven ML methods. We see that there is wide variability in the SimML predictor performances, with the XGB and KNN yielding the best performance. The analytical predictor also provides reasonable performance. Future work includes use of predictors in facility assignment problem.

REFERENCES

Fatma, N., S. Mohd, N. Mustafee, and V. Ramamohan. 2020. “Primary Healthcare Delivery Network Simulation using Stochastic Metamodels”. In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 818–829. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.