Forecasting daily incidence of Respiratory Symptoms: A Comparative Study on Time Series Models using OMOP-CDM in South Korea

<Min Ho An, MD>[1,2], <Min-Gyu Kim, MD>[1,2], <GyuBeom Hwang, MD>[1,2] ,
<ByungJin Choi, MD>[1,2], <Rae Woong Park, MD, Ph.D>[1]

[1] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea
[2]Department of Medical Sciences, Graduate School of Ajou University, Suwon, Republic of Korea

Background

With the outbreak of the COVID-19 pandemic, the significance of infectious disease surveillance and upsurge prediction has been emphasized. Several reports associated with prediction of respiratory infectious disease including COVID-19 is published.(1, 2) Respiratory infectious diseases like COVID-19 can disseminate rapidly, given the impossibility of restricting respiratory activities. To monitor disease spread, four distinct hospitals in South Korea recently began to collaborate to collect data using the Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) under the project named PHAROS (Platform for Harmonizing and Accessing Data in Real-time Infectious Disease Surveillance). During its nascent developmental stage in this project, we sought to compare two potent models, ARIMA(1) and Prophet(3), to predict the daily occurrence of respiratory symptoms. This study aims to assess each model's effectiveness and verify their accuracy in predicting the daily incidence of respiratory symptoms.

Methods

Patients visited or admitted to the emergency or infectious disease department presenting with symptoms including fever, dyspnea, or cough at Ajou University Hospital in South Korea were defined as respiratory symptom related visit. A total of 18,839 visits with respiratory symptoms were recorded from January 1, 2018, to December 31, 2021. The primary outcome in this study was the daily occurrence of respiratory symptoms classified above. To forecast this, we employed two models: ARIMA and Prophet. The total dataset was divided to train and test data, first allocating 80% towards the training set to build the model. The remaining 20% of the data was reserved as a test set to evaluate the model's predictive accuracy. All analyses were performed via Python v3.7.

Table 1. Summary of number of visits and respiratory symptoms in each year

	number of visits	fever	cough	dyspnea
2018	5523(29.3%)	3315(29.7%)	422(42.7%)	1786(26.7%)
2019	5563(29.5%)	3492(31.3%)	301(30.5%)	1770(26.5%)
2020	3775(20.0%)	2254(20.2%)	125(12.7%)	1396(20.9%)
2021	3978(21.1%)	2107(18.9%)	140(14.2%)	1731(25.9%)
Total	18839	11168	988	6683

Results

Table 1 reveals a marked decrease in visits since 2020. Both ARIMA (**Fig. 1**) and Prophet (**Fig. 2**) forecasts demonstrate similar outcomes, with most forecasted values lying within the 95% confidence interval for both models. Yet, the ARIMA model reported lower Mean Absolute Error (MAE) [2.66 vs 2.87] and Root Mean Squared Error (RMSE) [3.34 vs 13.10] than the Prophet model for test data (**Fig 1-2**). Intriguingly, the Prophet model better reflected the variance of the observed values than ARIMA, which primarily illustrated the downtrend with minimal variance.

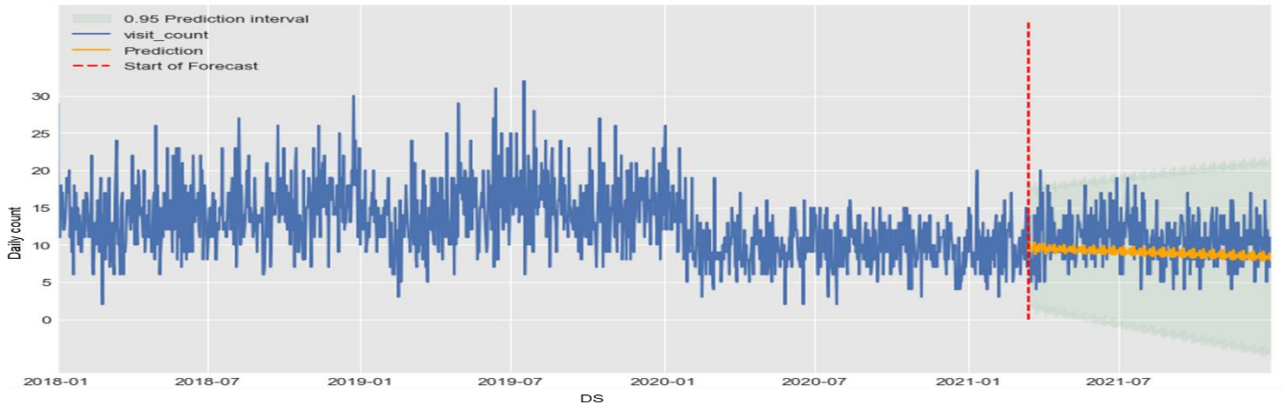

Figure 1. Daily count forecast using the ARIMA model. MAE: 2.66, RMSE: 3.34

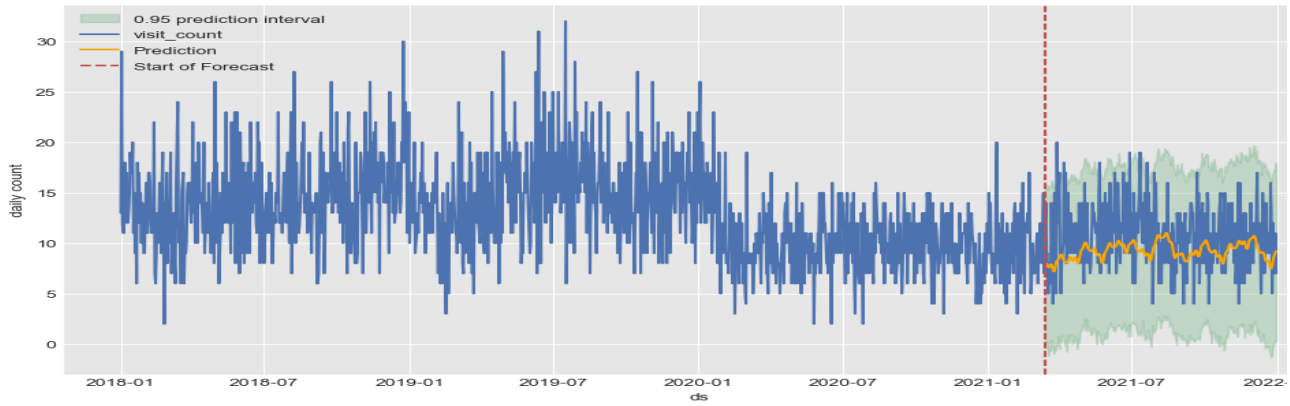

Figure 2. Daily count forecast using the prophet model. MAE: 2.87, RMSE: 13.10

Conclusions

In the task of predicting daily counts of respiratory symptoms in South Korea, the ARIMA and Prophet models mostly presented forecasts within a 95% confidence interval. Despite ARIMA's superior accuracy, denoted by a lower MAE and RMSE, the Prophet model offered a more realistic reflection of the data's variance. Therefore, model selection hinges on the study's specific objectives: ARIMA for numerical precision, and Prophet for discerning variance and trend changes. This study emphasizes the imperative of additional research to refine these models, enhancing infectious disease surveillance—a key component of healthcare preparedness in pandemic scenarios.

References

1. Somyanonthanakul R, Warin K, Amasiri W, Mairiang K, Mingmalairak C, Panichkitkosolkul W, et al. Forecasting COVID-19 cases using time series modeling and association rule mining. BMC Medical Research Methodology. 2022;22(1):281.
2. Krymova E, Béjar B, Thanou D, Sun T, Manetti E, Lee G, et al. Trend estimation and short-term forecasting of COVID-19 cases and deaths worldwide. Proceedings of the National Academy of Sciences. 2022;119(32):e2112656119.
3. Taylor SJ, Letham B. Forecasting at Scale. The American Statistician. 2018;72(1):37-45.