Predicting Air Ticket Demand using Deep Neural Networks

Kodai Imanaka and **Chiaki Sakama**
Wakayama University, Japan

IEEE BigData 2021

Background

- Nowadays airline tickets are mostly booked on the Internet and ticket sales change every moment. Airline companies use dynamic pricing to maximize their profits.
- Air ticket demands are also of particular interest for travel agencies who sell tickets to customers. It is crucial for agencies to predict market demand from the viewpoint of financial management.
- In addition to dynamic change in the booking status, however, the types of airline tickets sold by travel agencies are diverse, which makes it difficult to predict the sales of airline tickets.

Purpose

- We use deep neural networks to predict air ticket demand based on the past ticket sales.
- To this end, we introduce a new learning model that extends the LSTM by introducing exogenous variables for non-time series data.
- We evaluate on real data provided by a travel agency to predict air ticket demands of several flights by different airlines.
- We compare our prediction model with an existing time series model.

Dataset

- Data presents reservations of international flights arriving/departing the major airports of Japan.
- The number of flights is approx. 800 per day, and data contains the reservation status of each flight for the next 6 months on each day.
- We use data of 4 air routes: Tokyo-Paris, Osaka-Taipei, Osaka-Helsinki, and Nagoya-Honolulu, where each route has different airlines.
- Departure day is from 10/6/2018 to 30/9/2019 (approx. 16 months).

Data Attributes

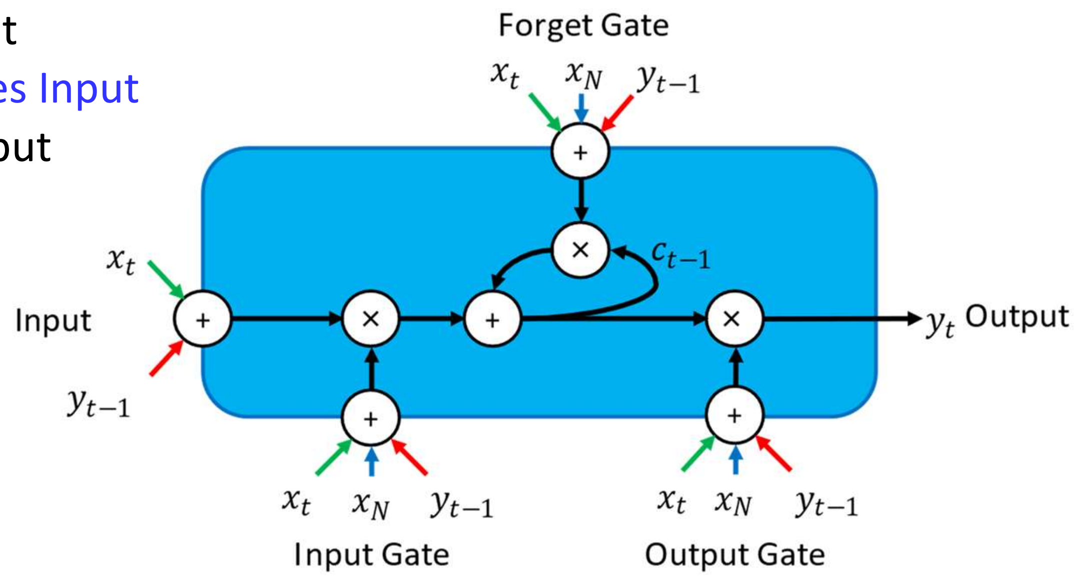
Attributes	Value	Data Type
outbound/inbound	0 or 1	non-time series
airline company	airline code	non-time series
departure date	Y/M/D	non-time series
departure time	0:00 - 23:59	non-time series
booking class	A-Z	non-time series
remaining seats	integer from 0 to 9	time series
day of the week	integer from 0 to 6	non-time series
holiday	0 or 1	non-time series

The fluctuation of the remaining number of tickets is expressed as time series data.

Learning Model

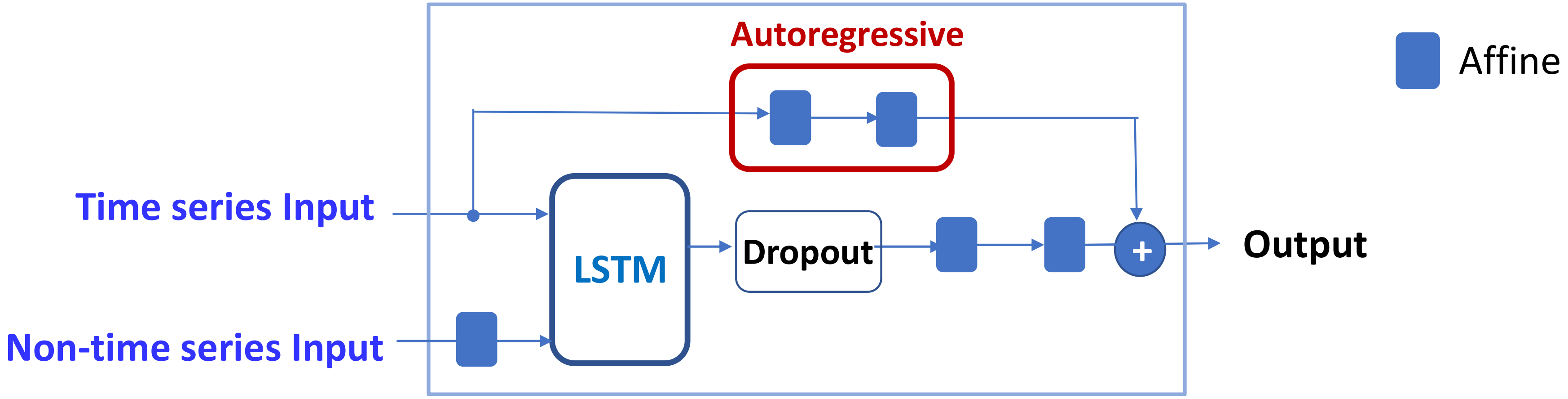
We introduce exogenous variables to the **Long Short-Term Memory** (**LSTM**) for handling non-time series data.

$y_{\text{t-1}}$: Recurrent Input
x_{N} : Non-Time Series Input
x_{t} : Time Series Input



LSTMX

- Time series input is fed into the autoregressive component.
- Output of LSTM is input into the Dropout layer.
- Output of LSTMX is obtained by combining the LSTM part and the autoregressive component.

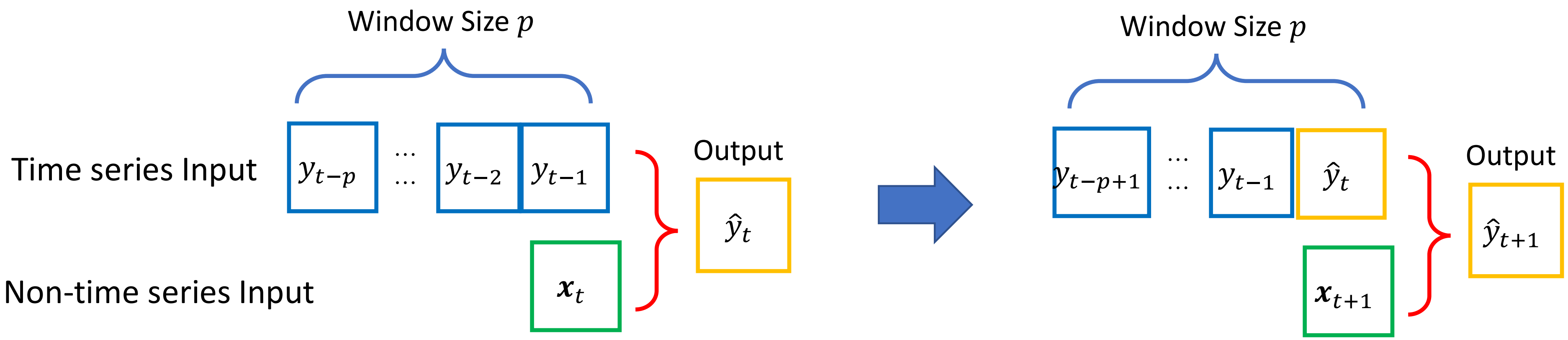


Learning Method

- A learning model is constructed for different booking classes for each flight in different routes.
- Sequences of data for 14, 28, 56 or 84 days are set as window sizes and fed into the LSTM.
- The number of remaining tickets of departure on the next day after a sequence is used as training data.
- Predict the number of remaining tickets on 14 days before the departure date.

Prediction Method

1. Using the latest time series data $y_{t-1}, \ldots, y_{t-p}$ with the window size *p* and non-time series data x_t, predict the number $\hat{y}_t$ of remaining tickets.
2. Incorporate the predicted value $\hat{y}_t$ into the time series input, then predict the number $\widehat{y_{t+1}}$ of remaining tickets on the next day.



Experimental Evaluation

- We compare prediction by LSTMX with SARIMAX that is one of the most popular techniques used to forecast a time series.
- We use RMSE (Root Mean Squared Error) to see prediction accuracy.
- Prediction is made for different booking classes of the following flights.

Air routes	#Airlines	#Flights	Business	Economy
Tokyo-Paris	3	8	40	40
Osaka-Taipei	3	20	78	100
Nagoya-Honolulu	2	4	20	12
Osaka-Helsinki	1	2	10	10

The number of booking classes in Business/Economy

Comparison of RMSE

- The value of RMSE represents the average of the difference between the real value and the predicted value of the number of remaining tickets. So smaller RMSE reflects greater accuracy.
- The following table represents the number of booking classes classified by the range of RMSE.

x means RMSE

		Business Class			Economy Class		
model	window size	x≦3	3<x≦4	4<x≦5	x≦3	3<x≦4	4<x≦5
LSTMX	14	38	46	38	24	32	42
	28	37	51	34	22	35	44
	56	38	48	38	24	34	42
	84	37	48	41	22	33	45
SARIMAX		38	30	26	18	30	39

Comparison of RMSE

- In business class, LSTMX and SARIMAX make prediction within RMSE≦3 in almost the same number of booking classes, while LSTMX makes prediction within RMSE≦4 or 5 in more booking classes than SARIMAX.
- In economy class, LSTMX makes prediction within RMSE≦3, 4 or 5 in more booking classes than SARIMAX.

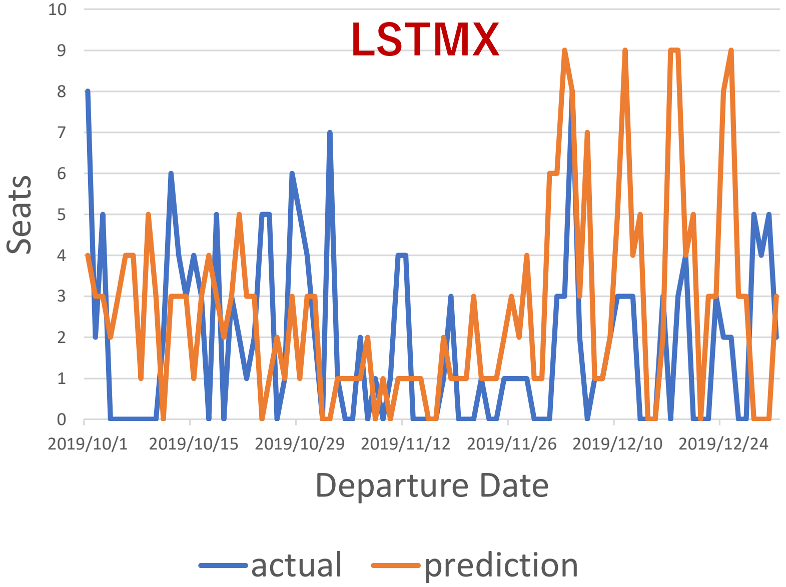
x means RMSE

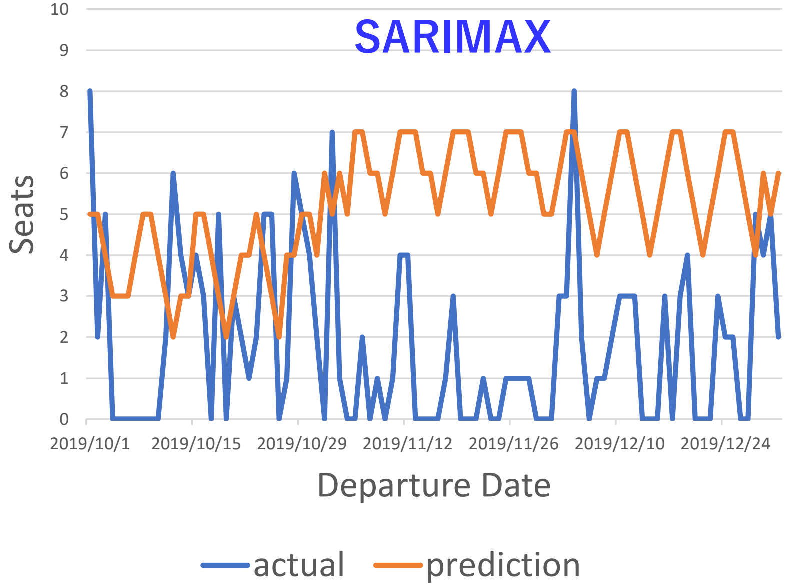
		Business Class			Economy Class		
model	window size	x≦3	3<x≦4	4<x≦5	x≦3	3<x≦4	4<x≦5
LSTMX	14	38	46	38	24	32	42
	28	37	51	34	22	35	44
	56	38	48	38	24	34	42
	84	37	48	41	22	33	45
SARIMAX		38	30	26	18	30	39

Prediction Results

LSTMX takes wider ranged values in prediction, and captures dynamics better than SARIMAX.

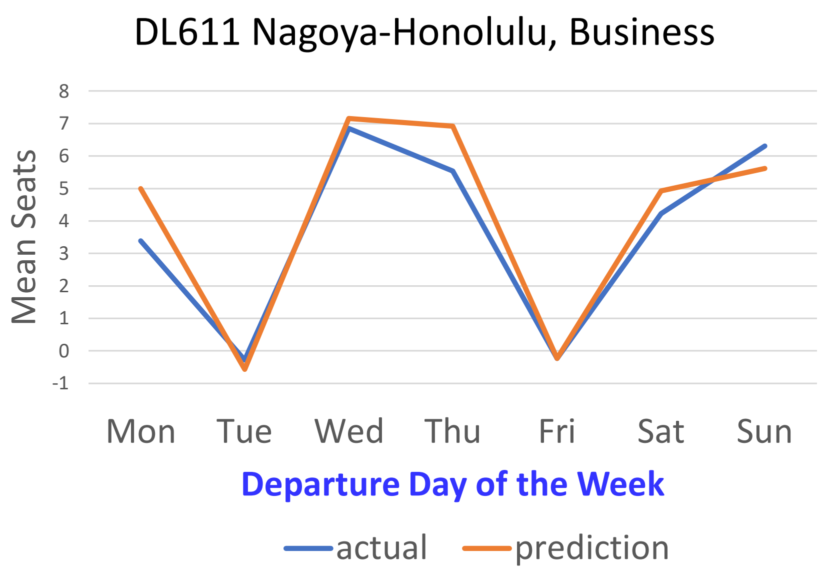
JL814 Osaka-Taipei, Business, Booking Class=C

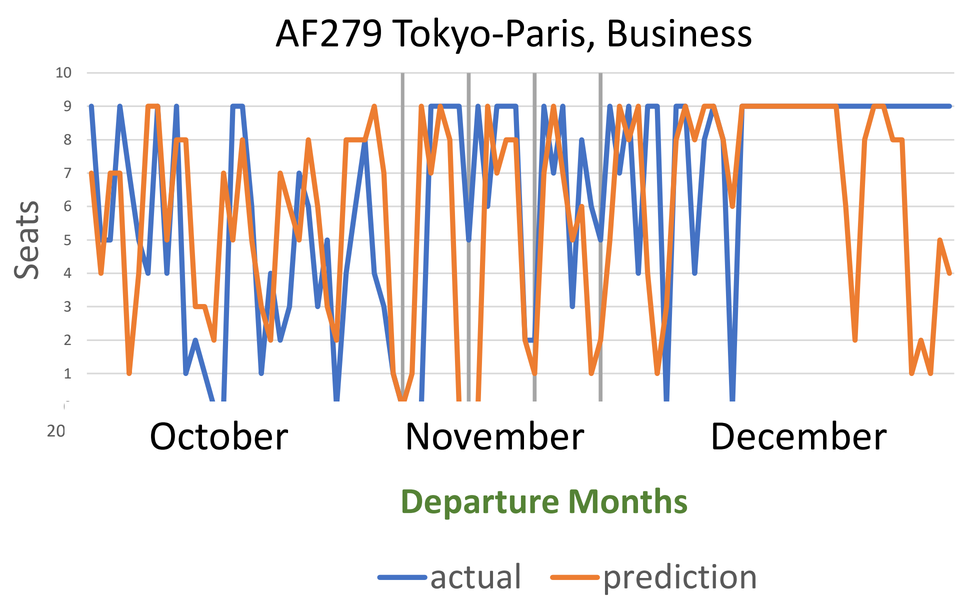




Prediction Results

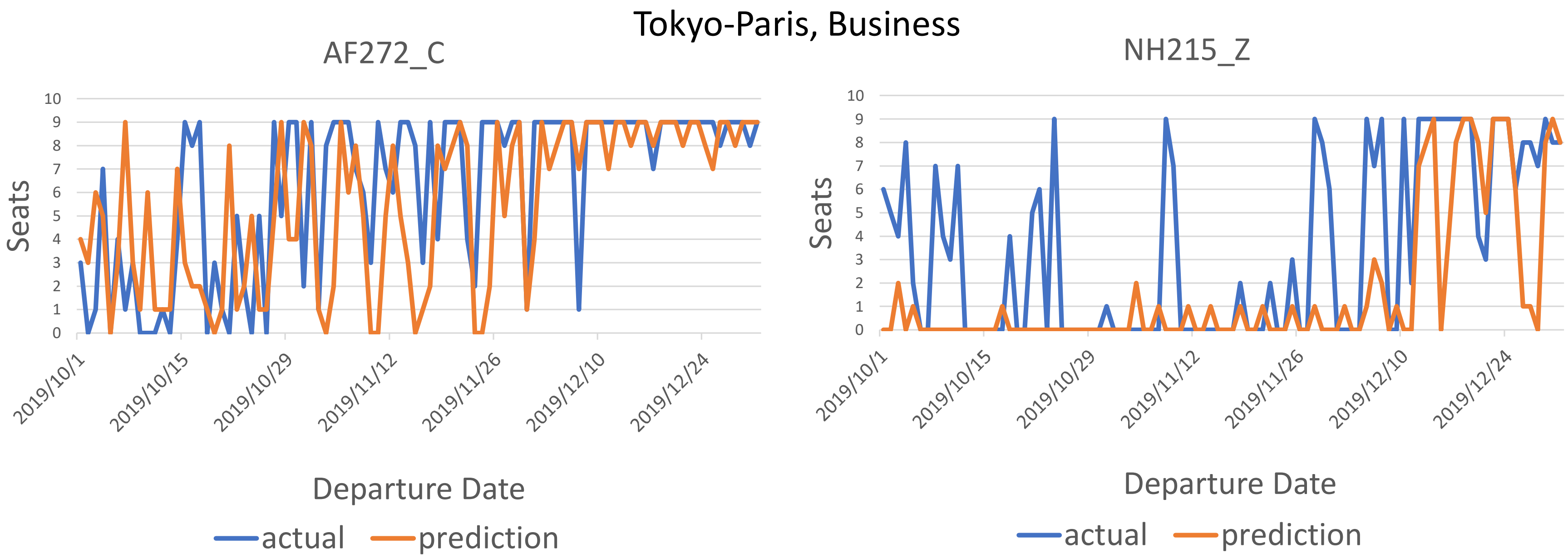
LSTMX captures weekly or monthly dynamics due to the mechanism of handling non-time series data.





Prediction Results

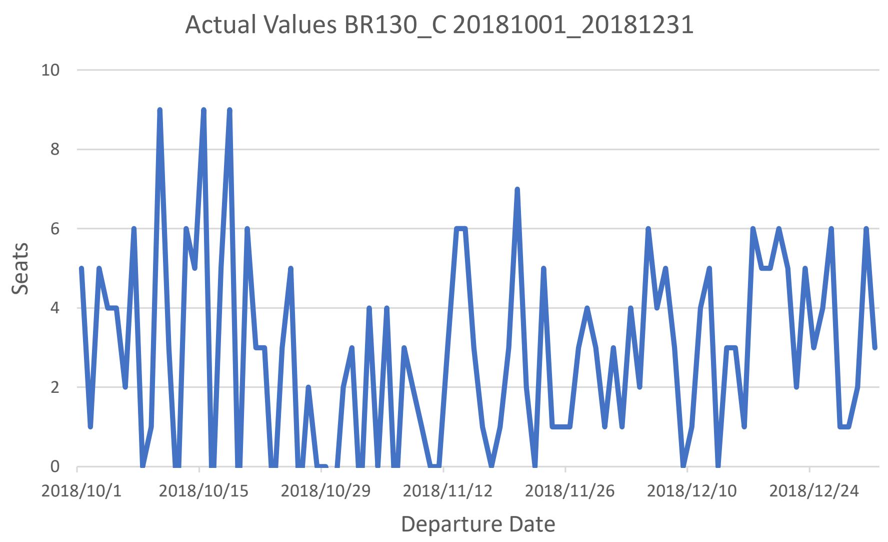
LSTMX captures different characteristics of different airlines on the same route.



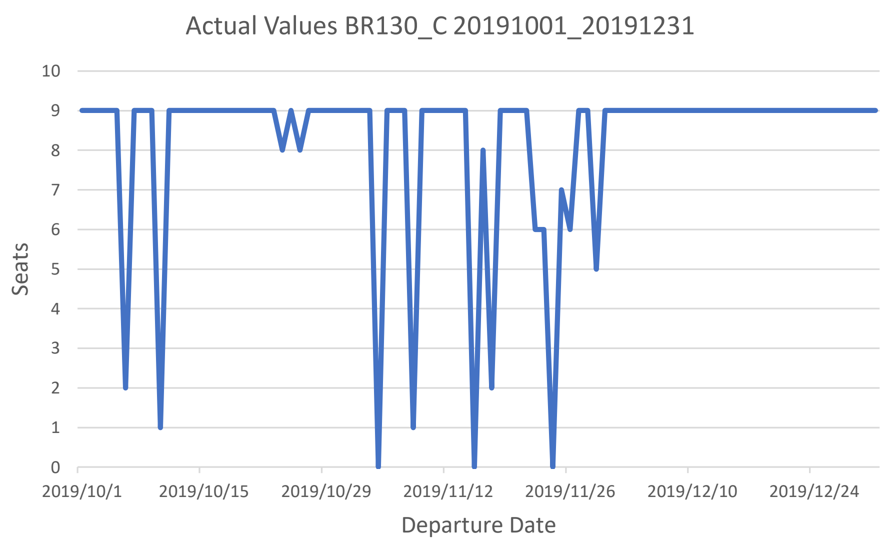
Prediction Results

There are cases where prediction fails significantly. One reason for this is that the same flight has different fluctuation in the same period.

BR130 Osaka-Taipei, Economy, 1/10/2018 - 24/12/2018



BR130 Osaka-Taipei, Economy, 1/10/2019 - 24/12/2019



Conclusion

- We apply a machine learning technique for predicting the number of remaining airline tickets using real data. A new learning model LSTMX is introduced to handle non-time series data as well as time series data.
- Experimental results show that the LSTMX successfully predicts weekly/monthly changes and distinguishes different airlines on the same route.
- To improve the prediction accuracy, it is necessary to increase the amount of data over several years.