A Novel Model Predicting Booking Curve Based on OTA Search and Transaction Data

Quanwu Xiao & Siyuan Zhao

Ctrip & Trip.com

Revenue Management in Airlines

Simply put, we use market demands to decide how to price tickets.

High demand → **High** load factor → **High** ticket price

Low demand → **Low** load factor → **Low** ticket price

Disadvantages of airline companies direct RM

- **Their ways to get information**
  - Historical information
  - Lack of information about other carriers
  - Manual price comparison
- **Faults of pricing by man**
  - Cost of employee training
  - Human error
  - More expensive than machine

Revenue Management from OTA

Ctrip 携程 Trip.com

OTA understands demand better !

High Market Share

With a quite high market share in China (aviation), OTAs can represent market trends.

High Search Volume

Customers usually search for tickets several days before booking. Customers who don’t buy ticket on OTAs also search on them. So they have plenty of data about future market demand.

Cross-airline Booking

Customers can choose different carriers on OTAs.
Also they provide a convenient cross-airline transfer/RT booking process.

Sufficient Price Information

OTAs naturally have prices of all carriers so that customer selection model can be built and real-time price comparison can be performed.

Load Factor and Booking Curve on Route

Target selection

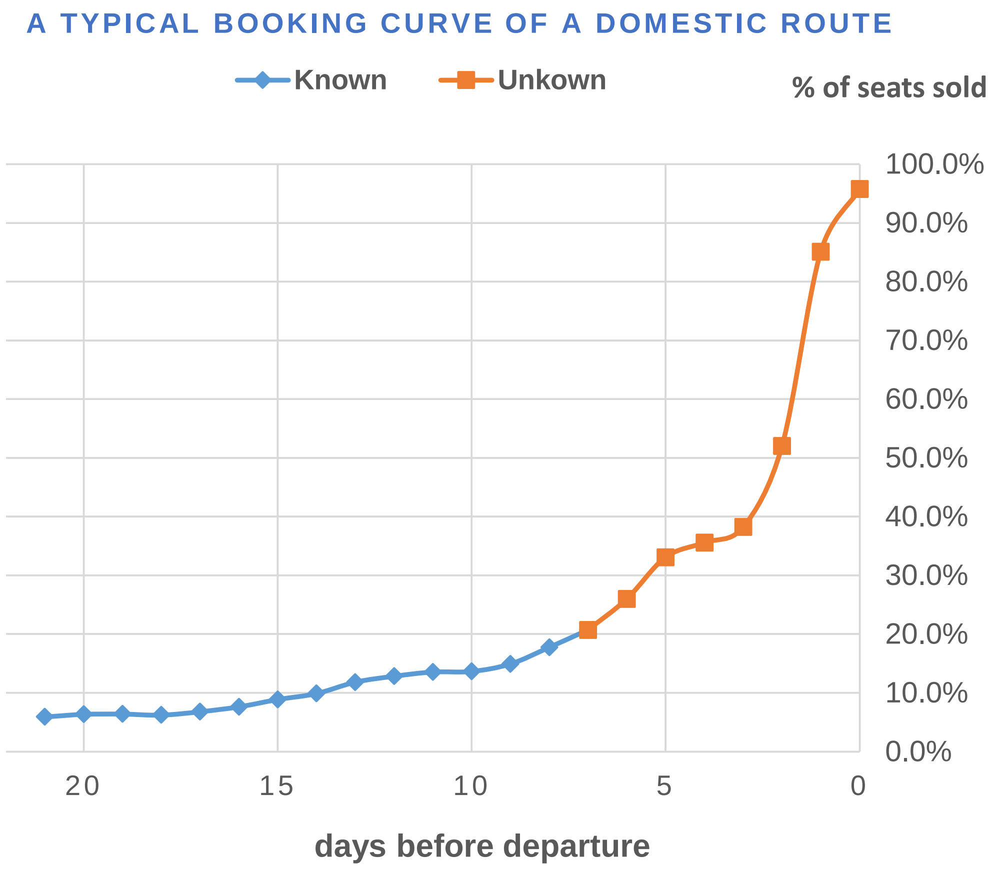
- Flight Load Factor : unstable, easily affected by human operation
- Route Load Factor : more stable, less affected by a single carrier

$\mathrm{LF_t}$ stands for Load Factor, which is the percentage of seats sold of a route t days before departure:

$$LF_{\mathrm{t}} = \frac{\sum_f S_{f,t}}{\sum_f C_f}$$

where f represents the flight of a particular route, and $LF_t \in [0, 1]$

- LF_t: **route passenger load factor** t days before departure.
- $S_{f,t}$: **number of seats sold** of flight f t days before departure
- C_f: **capacity** of flight f, assumed to be fixed



Machine Learning Approach

Key Features:

1. Search popularity
2. Known sales progress
3. Business traveler ratio
4. Holidays
5. Historical passenger load factor progress
6. Capacity
7. Flight departure time distribution
8. LCC ratio
9. Seasonal trends

……

X

$$y \sim \hat{y} = f(x)$$

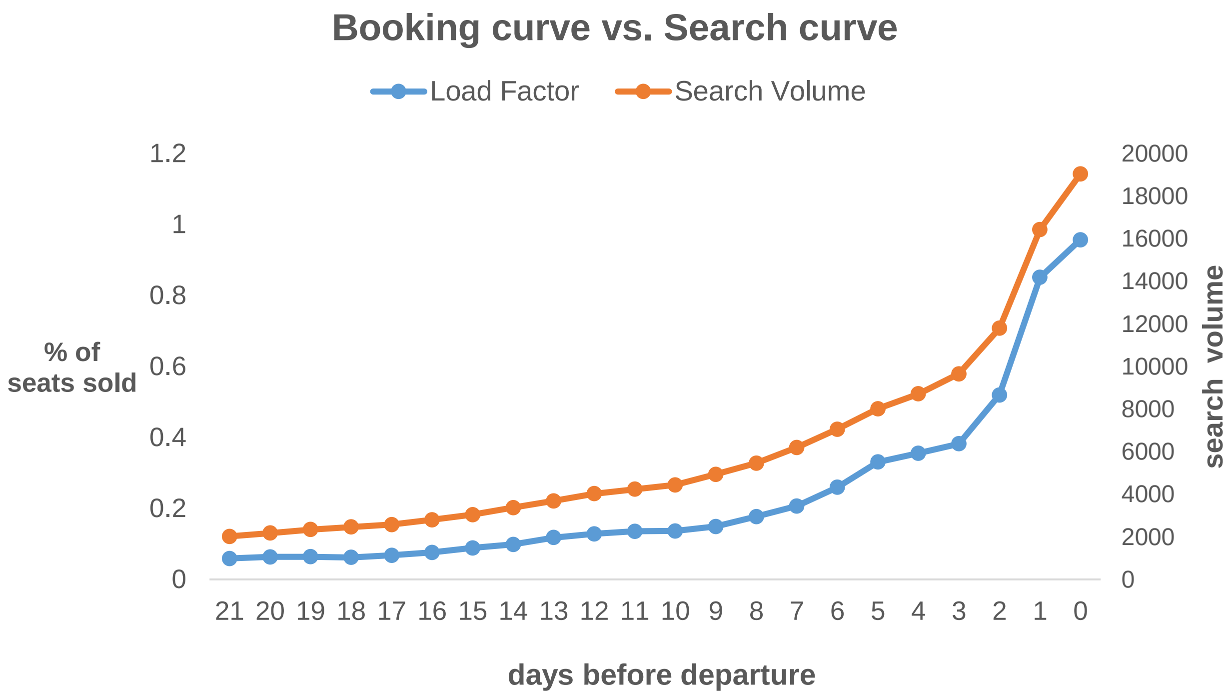
Y

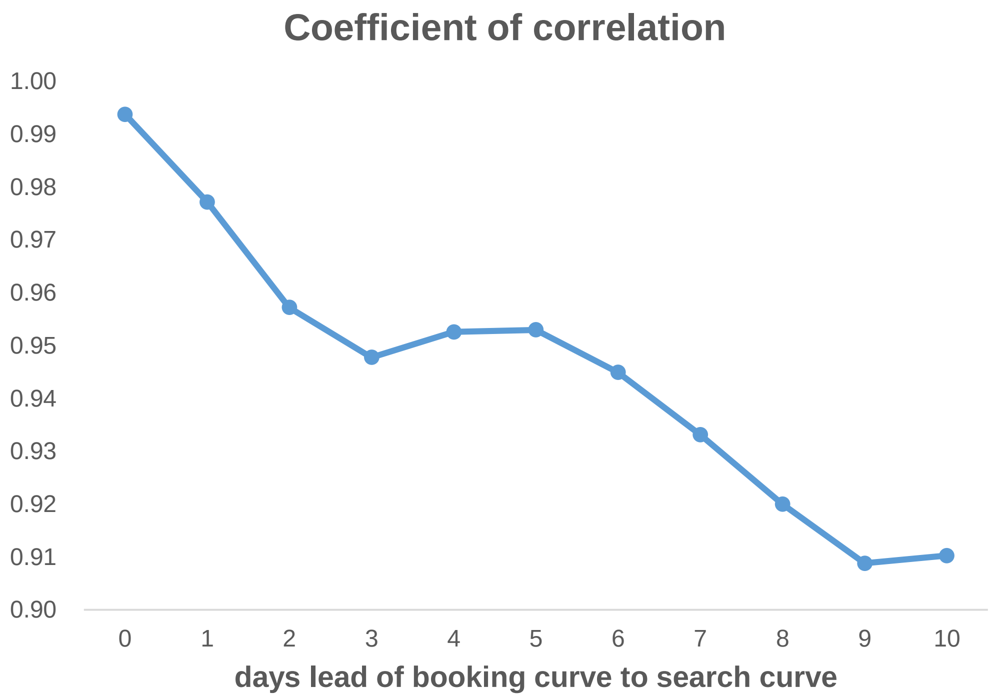
Future booking curve (% of seats sold) of a route

Here we do not consider origin-destination control based revenue management.

Search Volume

- People tend to search for prices before booking
- Changes in search volume reflect market changes
- Correlation is above 0.9 when we align the booking curve with the search curve within 10 days in advance.

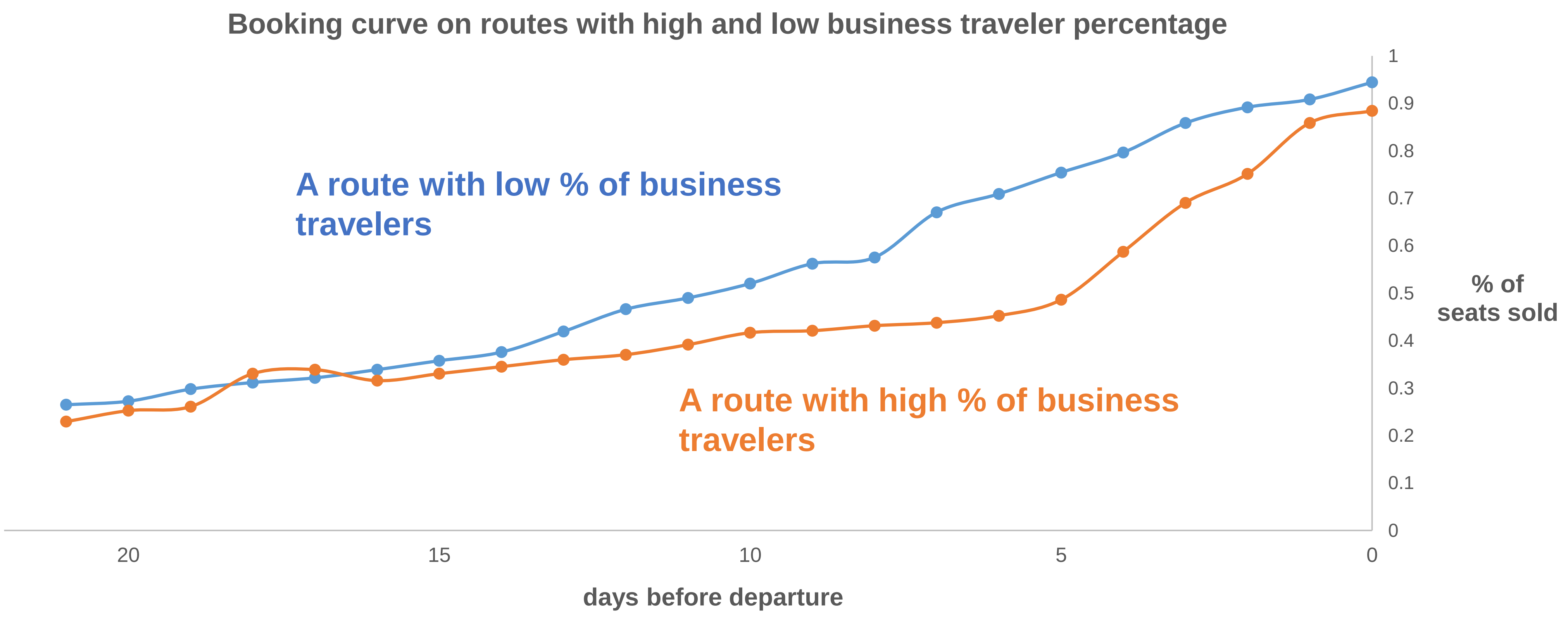




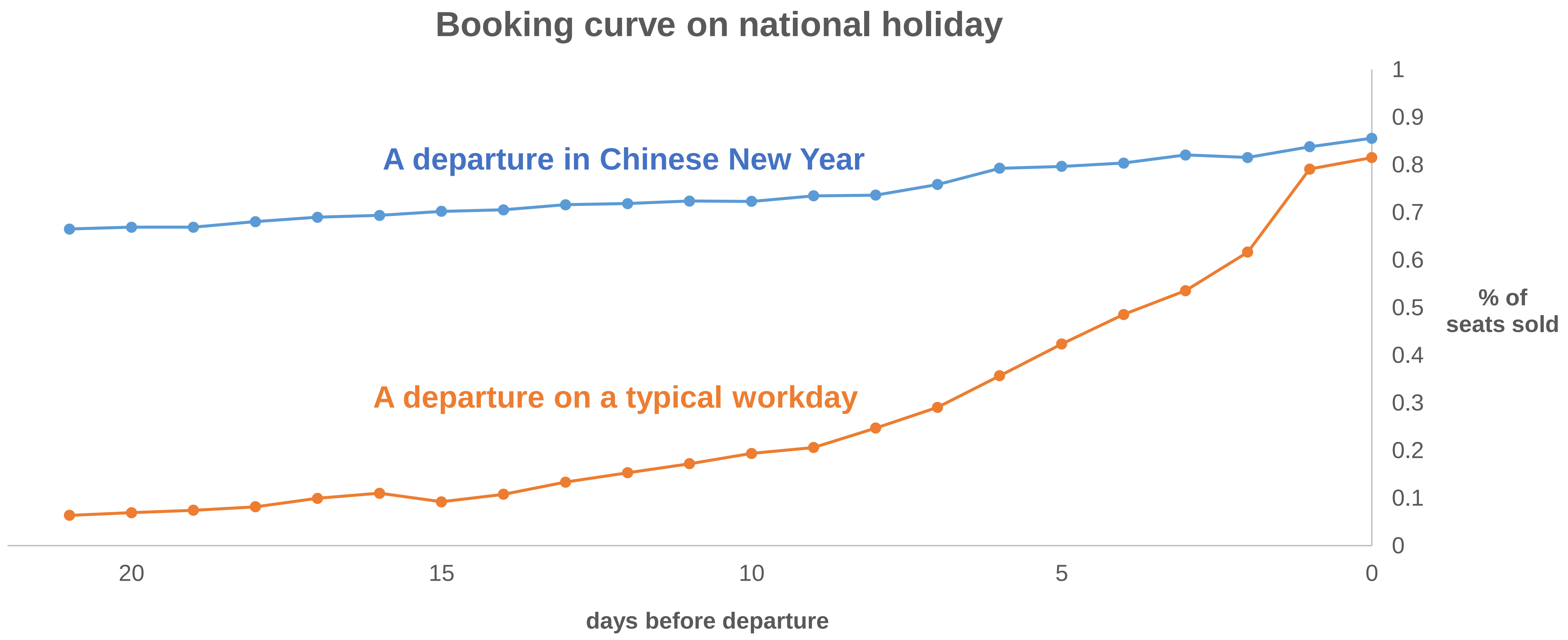
Percentage of Business Travelers

Definition of **business travelers**

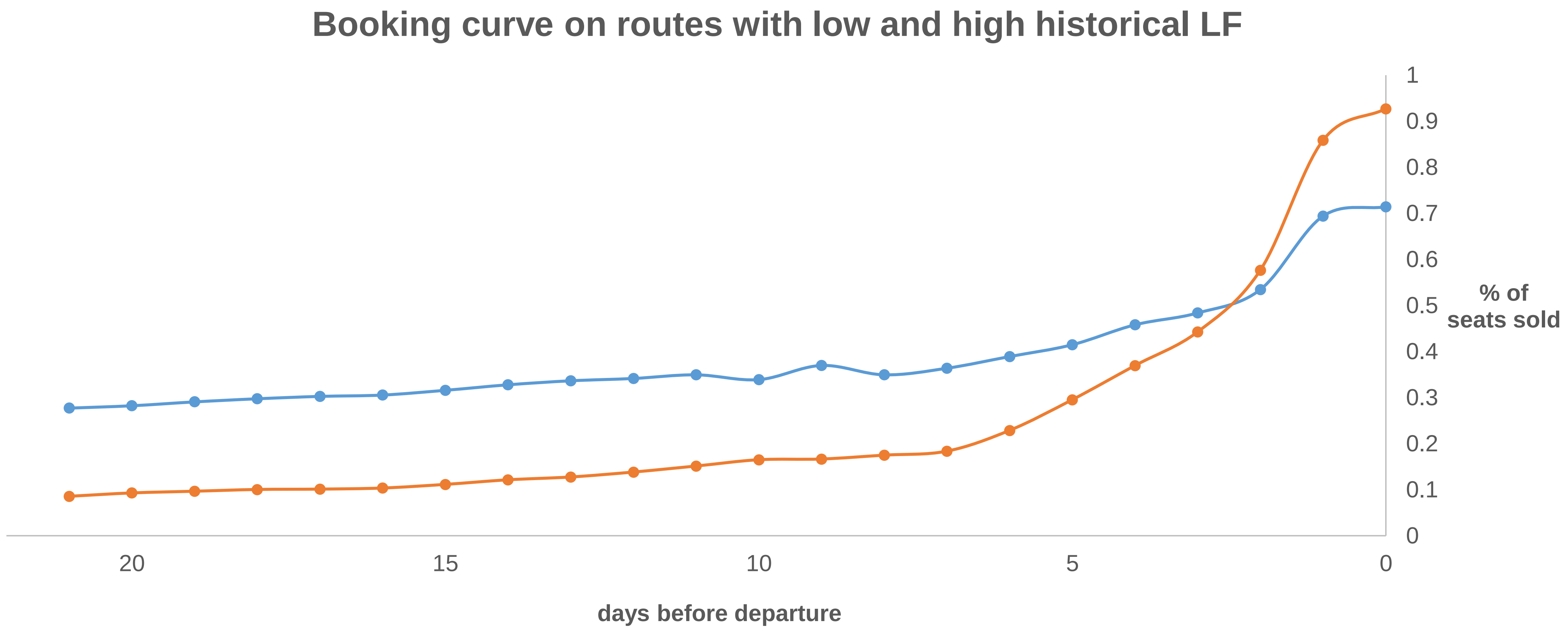
- Need reimbursement
- Book through TMC channel
- Other clues



Holiday


Booking curve on national holiday
A departure in Chinese New Year
A departure on a typical workday
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
% of
seats sold
20
15
10
5
0
days before departure

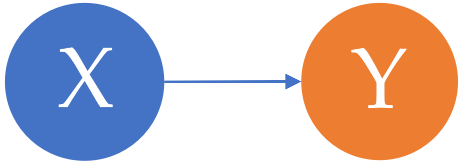
Historical Passenger Load Factor



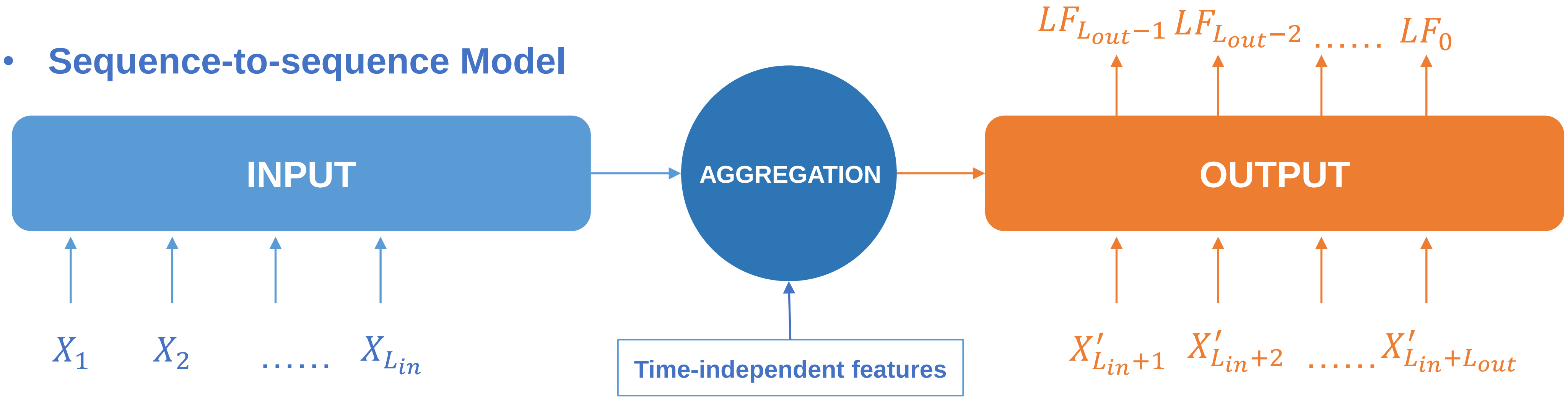
Sequence-to-sequence Model

Ctrip 携程 Trip.com

- **Traditional Model**



- **Sequence-to-sequence Model**



- L_{in}: Input sequence length
- L_{out}: Output sequence length
- X_t: Input time-dependent features
- X'_t: Output time-dependent features
- LF_t: Passenger load factor t days before departure

Reference： Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, pages 3104–3112.

Advantages of Sequence-to-sequence Model

Ctrip 携程 Trip.com

Traditional TS Model

Not easy to integrate time-independent features

Traditional ML Model

Estimate with 14 or more models

Sequence-to-sequence Model

14 models are replaced by 1

1. Update predictions daily

2. Receive variable input & output sequence length

3. Model different routes together

4. Make prediction based on previous predictions

5. Utilize different types of feature

time-dependent & time-independent
categorical & numerical
past & future information

Model Performance

Mean Absolute Error (MAE)

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \widehat{y}_t|$$

- where $(y - \hat{y})$ represents error
- over **300+** domestic routes
- departing in the next **one** month

MAE	Gradient-Boost	Seq2seq
With OTA data	4.61%	**4.39%**
Without OTA data	-	5.10%

Days before Departure / Days making prediction	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0															2.21%
1														3.34%	3.87%
2													2.94%	4.49%	4.60%
3												2.51%	4.22%	4.90%	4.90%
4											2.22%	3.63%	4.75%	5.20%	5.14%
5										2.01%	3.23%	4.31%	5.14%	5.38%	5.26%
6									1.76%	2.94%	3.93%	4.78%	5.44%	5.52%	5.38%
7								1.79%	2.73%	3.69%	4.47%	5.20%	5.73%	5.66%	5.46%
8							1.67%	2.70%	3.48%	4.26%	4.91%	5.54%	5.95%	5.78%	5.52%
9						1.48%	2.44%	3.32%	3.97%	4.66%	5.22%	5.75%	6.15%	5.88%	5.57%
10					1.41%	2.26%	3.03%	3.80%	4.36%	4.96%	5.46%	5.93%	6.29%	5.94%	5.61%
11				1.33%	2.15%	2.85%	3.53%	4.19%	4.70%	5.25%	5.68%	6.10%	6.44%	6.01%	5.64%
12			1.24%	2.00%	2.73%	3.33%	3.93%	4.54%	5.01%	5.53%	5.92%	6.28%	6.62%	6.09%	5.69%
13		1.15%	1.79%	2.49%	3.14%	3.67%	4.26%	4.82%	5.26%	5.74%	6.12%	6.46%	6.80%	6.15%	5.74%
14	1.13%	1.70%	2.31%	2.92%	3.51%	4.00%	4.54%	5.05%	5.46%	5.91%	6.30%	6.63%	6.96%	6.22%	5.80%

Case: Predict Load Factor For Pricing

MAEs of all 6 routes are **within 5%**, even 15 days before departure.

- For baseline approach, the carrier uses historical load factor from last year as an estimate for this year.

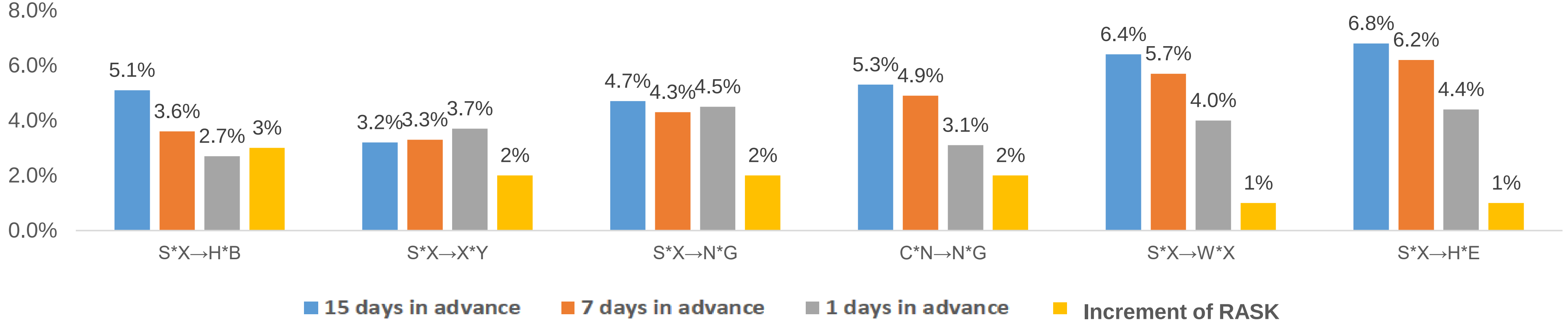
Compared with either other routes of this particular carrier or the same route of competitive carriers, **revenue** of all 6 routes has **increased**, with the average rate of **2%.**

- The strategy to use the load factor prediction for this particular carrier is illustrated in the right chart.

Pricing Strategy

	High Predicted Final LF	Low Predicted Final LF
Faster Sales than market	Increase	Increase/decrease slightly
Slower sales than market	Increase/decrease slightly	decrease

Prediction Error & Performance



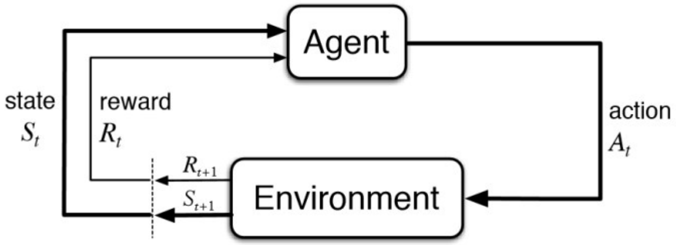
Future Improvement

Combine the predicted booking curves with the following technologies.

Customer choice model

Dominating Price		￥2790	
Follower Price	￥2340	￥2490	￥2700
Booking Probability	82.8%	81.2%	64.1%
Expected Revenue	￥1939	***￥2023***	￥1730

Reinforcement learning and simulator



We are open to academic and industry communities for collaboration.

Thanks!