Forecasting the Demand of Library Booking Rooms to Promote and Manage Co-working Space Efficiently

BAFT Group 1

Shu-Ming Yeh, Pei-Rong Chen

Pi-Ruei Jiang, I-ting Wei

01/14/2015

Business Problem and Background:

We are going to found a company turning the redundant space in campus into appealing co-working spaces for students. Since we have no any history record for predicting future demand of discussion room. The existing booking data of discussion room of library can, to some extent, reflect the pattern of real needs. Therefore, the project is going to forecast the demand of discussion room as accurate as possible to help our company better prepare and allocate resources.

Forecasting Goal:

Predict future weekly booking in the following semester

Implement Detail and Results

We have only two semesters data available to predict. We identify that the main users are from college of engineering (CE) and technology management (CTM). Therefore, we build two models to predict each of them. After capturing the pattern of two colleges. We try to further predict potential customer component of the CTM, hence, we build three models to predict three main departments of CTM, finding out that only prediction of department quantitative finance is relative reliable.

Recommendation and Limitation

Based on the results, we would suggest that we have to prepare the maximum of capacity of seat is about 200 (combining two colleges results), and we would suggest the regularly maintenance should be performed on the week 63(detail showing in the figure in report). However, the goal of using history record to customize some activities and promotion campaigns for specific department may have some concerns. Our suggestion is to wait for more data collecting and do roll-forward forecasting. And I also suggest the future work should devote to building one reliable model to do all forecasting work to reduce labor work.

- **Problem description**

Facing inconvenience in library; for example, no food, no drinking, no enough space to discussed, and time limited for discussion rooms' usage. The idea of transform redundant space left in campus into co-working spaces come out. Since the record of library rooms can actually reflects the pattern of need of co-working space, this project is going to forecast the demand to run a co-working spaces.

Business goal

Purpose: Using Forecasting to analyze the data to deliver the advertisement to the right people at the right time and right place.

i. Stakeholder: Ourselves, an imaginary co-working spaces company.

ii. Opportunity: Better allocate advertisement resource, arrange space utilization and promote campaign.

Forecasting goal

i. Forecasting objective: The number of people booking spaces.

ii. Forecasting horizon: 18 weeks (a semester) (forward-looking goal)

iii. Forecasting usage: With the forecasting result, when and who will be the high volume user for discussion room will be known. This can help us making different adverting strategy to every college and departments, attracting the users of library and transforming them to be our customer.

- **Data description**

The dataset is come from Library of NTHU. It is the daily amount of booking rooms in Library. Time period for the dataset is from March 25th 2013 to October 29th 2014. The total amount of data are 121,774 records. The sample of raw data and the meaning of each column is shown below.

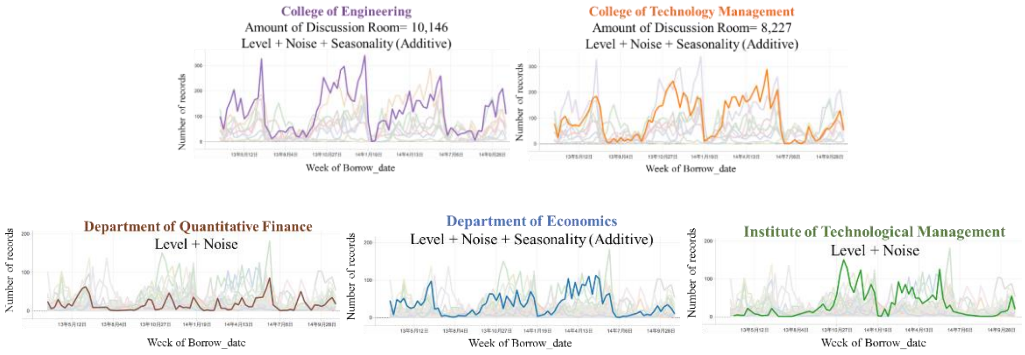
Type_space	Department	Borrow_date	Borrow_time	college	undergraduate
1	455	2013/3/25	16:01	4	2
2	820	2013/3/25	16:04	1	1
2	820	2013/3/25	16:04	1	1
2	810	2013/3/25	16:04	1	1
1	420	2013/3/25	16:06	4	1
1	745	2013/3/25	16:09	2	2
1	745	2013/3/25	16:15	2	2
2	230	2013/3/25	16:16	6	1
1	420	2013/3/25	16:20	4	1
1	311	2013/3/25	16:20	5	1

- Type space: 1 is individual study room; 2 is discussion room
- Department: 56 different departments including mba students and school staffs
- Borrow_data: 2013/03/25~2014/10/29, except for the non-service day
- Borrow_time: from 8am to 10pm, which is the opening time for library
- College: 9 colleges (Science, Engineering, Electrical Engineering and Computer Science,

Commission of General Education, Humanities and Social Sciences, Life Science, Nuclear Science, Technology Management and school staffs)

- Undergraduate: 1 for undergraduate students and 2 for graduate students

The library collected every record including user open or close the door, turn on and off the air conditioner and lamp. Since we cannot distinguish whether the record represents the room is occupied or the user just leave for bathroom, we cannot use the data for individual study room.



- **Data preparation**

For the data preparation, we do three changes on the raw data.

i. From the visualizing time plot, we observed that the period for seasonality are semester dates, so we decide to remove summer and winter vacations to do a better forecast.

ii. We know that students need to enter discussion room within 15 minutes of the booking time. For example, time form 08:15~08:30 and 08:45~09:00 are not a correct time for users to active the discussion room. We used this rule to distinguish whether the record represents the room is occupied or the user just leave for bathroom. Then, we delete the time which not allow users entering.

iii. In the beginning, we found linear regression with 144 dummies result in a good forecasting. But we can no longer use dummies over 100 anymore, because of the software limitation, the latest version of Xlminer does not allow us to do so. Instead of 144 dummies, we aggregate the daily data into weekly data and create 18 dummies for every week in a semester.

The table below is the chart after we remove vacations and incorrect records, and the right is the chart we aggregate into weekly data and create 18 dummies through Microsoft Office Excel's PivotTable report.

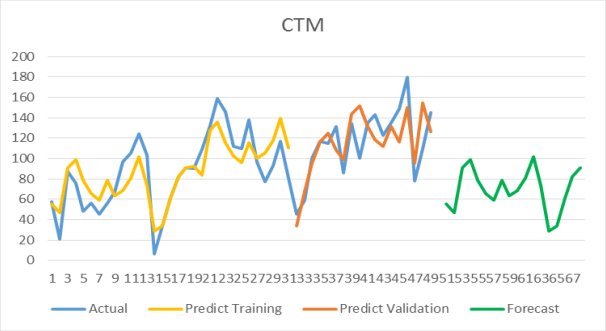
TYPE_SPACE	Department	Borrow_date	Borrow_time	college	undergraduate
2	331	2013/3/25	18:30:18	5	1
2	311	2013/3/25	19:30:14	5	1
2	331	2013/3/25	19:37:18	5	1
2	340	2013/3/25	19:37:43	5	1
2	331	2013/3/25	19:38:02	5	1
2	311	2013/3/25	20:13:25	5	1
2	331	2013/3/25	20:41:10	5	1
2	320	2013/3/25	20:59:16	5	1
2	300	2013/3/25	21:03:17	5	1
2	311	2013/3/25	21:05:20	5	1
2	331	2013/3/25	21:05:37	5	1
2	311	2013/3/25	21:10:46	5	1
2	300	2013/3/25	21:22:20	5	1
2	311	2013/3/26	10:32:05	5	1
2	311	2013/3/26	10:32:38	5	1
2	311	2013/3/26	10:39:08	5	1
2	311	2013/3/26	11:47:39	5	1

t	ce weekly aggregate	dummy
1	88	6
2	38	7
3	93	8
4	98	9
5	146	10
6	73	11
7	102	12
8	64	13
9	83	14
10	106	15
11	96	16
12	140	17
13	159	18
14	8	1
15	34	2
16	59	3
17	71	4
18	145	5

- **Forecasting solution& Time plots**

i. College of Technology management

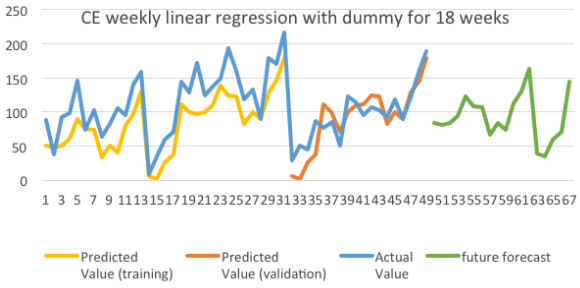
Methods: Multiple Linear Regression **<Evaluation 1>**



Validation Data Scoring		Seasonal-Naïve	
RMSE	Average Error	RMSE	Average Error
22.579	0.66667	47.7837	34.8333

ii. College of Engineering

Methods: Multiple Linear Regression **<Evaluation 2>**



Validation Data Scoring		Seasonal-Naïve	
RMSE	Average Error	RMSE	Average Error
23.6746	6.1204	59.28509	36.61111

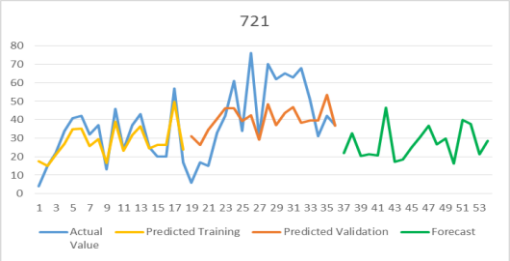
In the first two series , College of Technology Management (CTM) & College of engineering. We can see that there are seasonalties and a little trends in the series, hence, we chose multiple linear regression to do our forecast. Another advantage for us to use MLR is that we can quickly and easily add variables into our model if there are possible events held. For example : school's sports day and graduation ceremony.

And we can tell that our outcomes is much better than our benchmark, seasonal naïve forecast.

The following charts include actual , training, validation and forecast period

iii. Department of Economics

Methods: Neural Network **<Evaluation 3>**

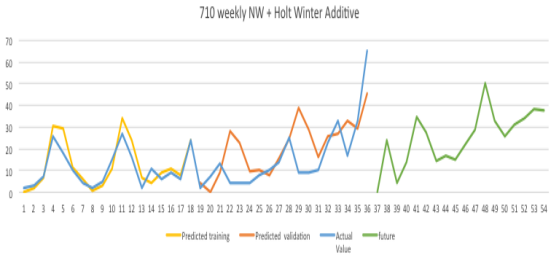


Validation Data Scoring		Seasonal Naïve	
RMSE	Average Error	RMSE	Average Error
12.4149	4.60069	22.3532	15.2222

The third time series is Department of Economics. As we can barely see trend and seasonality inside these series, we use neural network and multiple linear regression to do the forecast. The outcome shows that neural network has lower RMSE. Hence, we chose neural network as our forecasting method.

iv. Department of Quantitative Finance

Methods: Neural Network + Holt-Winters Additive **<Evaluation 4>**

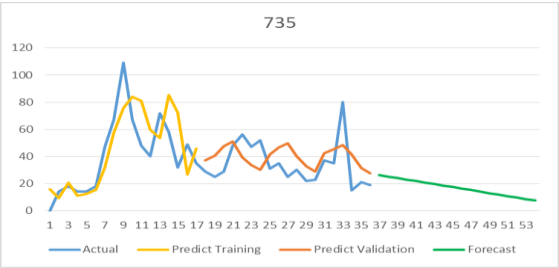


Validation Data Scoring		Seasonal Naïve	
RMSE	Average Error	RMSE	Average Error
16.075264	4.89929	34.9118	-5.6111

However, in the series of department of quantitative finance, we can't capture the future upward trend. As a result, we ensemble Holt-Winters additive to capture the upward trend. We successful got a better result and capture the upward trend. The RMSE is lower than the naïve forecast.

v. Institute of Technology Management

Methods: Holt-Winters Additive + Holt-Winters Multiplicative **<Evaluaiotn 5>**



Validation Data Scoring		Seasonal Naïve	
RMSE	Average Error	RMSE	Average Error
16.075264	4.89929	34.9118	-5.6111

The last time series is institute of technology management. We decide to hack a forecast better than seasonal naive. We surprisingly found that holt-winters additive's result is on average higher than actual series, and the holt-winters multiplicative's result is on average lower than actual series. As a result, we just ensemble these two holt-winters method together! And we got a better forecast successfully.

- **Conclusion**

In this study, we find out that our model can capture most volume of users by well forecasting college data. This is a quietly exciting result. However, there's still space for us to improve forecasting of departments' volume. Because one of our forecasting goal is to better allocate our resources including advertising, holding theme party, and preparing for other department-dependent activities. In summary, we achieve part of our goal to capture most of the volume, but we have to collect more data to build a more reliable model.

i. Advantage

For example, using our MLR model to predict future volume of CTM, we would have better accuracy (14% more accurate) than using naïve or seasonal naïve forecast (forecast tomorrow using today's data or forecast next Monday using data of this Monday).

ii. Recommendation

We suggest analyst of the future to collect more data and do rolling-forward forecast to get relative steady result and improve the accuracy. And, we also give an advice that future forecasting for different departments should try to use one universal model, or at most two models to predict. Otherwise, calibrating parameters of model or assembling two models would cost a lot of labor work

iii. Limitation

Because we built our model based on the data of "newborn" library. It indicates chaos in the first semester. Therefore, we can only use one of the remaining semester to train our model and use the other to validate. The forecasting result would be improve by keep collecting coming data.