

NUS Business School Honors Dissertation

Peng Seng Ang

You may put a date here

Abstract

This paper studies how we can use various spatial-temporal time series model to better predict demand across different locations.

1 Introduction

Having an accurate forecast of delivery demand for food service providers would help them more effectively and efficiently assign orders to drivers to improve the overall delivery time. Currently, most Autoregressive (AR) or Autoregressive Integrated Moving Average (ARIMA) models only consider temporal features when predicting demand. However, we believe including spatial features between the data points might improve forecast accuracy. This paper would focus on and explore models that include both spatial and temporal features to improve forecast accuracy.

2 Data

The data source used was an operational dataset from a food delivery service provider from Shanghai that includes delivery information for a 2-month period (excluding weekends) in 2015. The provider only provides delivery service for 90 minutes during lunchtime and the dataset has split the data into 15-minute time periods, and as such, each day would only consists of demand data for 6 time periods. Hence, our dataset has 839 locations with demand data for 204 time periods in total.

2.1 Exploratory Analysis

We can see.....

3 Baseline Model

In this section, we would build a simple baseline model. Following which, we would try other different spatial temporal time series models and compare the results to the baseline model. The main metric used for comparison would be Mean Squared Forecast Error (MSFE).

3.1 Train-Test Split

Firstly, we would split the dataset into training and test set by considering the first 27 days as the training set and the next 7 days as the test set. Our training set would then have 162 demand data for each location and test set would have 42 demand data for each location.

3.2 Baseline Model

A baseline model was first built by building a simple ARIMA model on each of the locations individually. Auto-arma function from Python was used to implement this. The MSFE for this baseline model on the 42 locations is 58.80.

Assumption 1. *It is because...*

Assumption 2. *It is generally true...*

4 VAR Model

Assumption 3. *It is because...*

Assumption 4. *It is generally true...*

4.1 Results

The results from the VAR model implemented by BigVAR using R gives a MSFE of 50.05 on the 42 locations.

5 GLM Model

Assumption 5. *It is because...*

Assumption 6. *It is generally true...*

5.1 Insights and Implementation

Any findings from the results? If there are any benefits or issues in implementing the proposed model...

6 Conclusion

Conclude your efforts and main findings.

7 Appendix

Append extra plots, graphs, analysis, etc.