CAPSTONE MILESTONE REPORT

Project:

Prediction of Vessel Arrival Time using Automatic Identification System (AIS) satellite data

Background:

Shipping companies are worried about potential port congestions or delays brought by new carrier alliance reshuffle in April 2017. And their worries become a reality in recent months. Shanghai port in China has hit port delays up to 1-2 days due to high volume of containers coming in thru Yangshan port aside from being the busiest port in the world.

Problem:

This project aims to

1. Improve the estimation of arrival time of vessels from an origin port to destination port using machine learning.
2. Investigate patterns of vessel routes at sea
3. Determine on how to detect or measure seasonality as a factor in estimating arrival time

Target Client:

The author’s current company is a target client for this project and also other logistics companies interested in improving their inventory planning and cost analysis.

Data:

Automatic identification system (AIS) data is a tracking system used to avoid collision on ships and other water transports. This system is mandatory for ships with gross tonnage of 300 or more and for all types of passenger ships.

Important data from AIS includes vessel id, position (latitude, longitude), speed, heading, course, date/time and destination.

Data Wrangling:

The data on destination is a free text and does not follow a specific rule. When this data is not properly cleaned, the study will give a different result since the whole dataset for a given voyage will be incomplete. Below are the reasons of why the data on destination needs a careful cleaning.

1. Format by United Nations (UN) location codes. For example; SG SIN or SGSIN is Singapore
2. Country port format like JP YKK SN which means Japan Yokohoma bound to Senegal
3. Origin port and destination port format like HKHKG > CNYTN is Hong Kong to China, Yantian
4. Spaces in between like ZHANG JIA GANG is ZHANGJIAGANG
5. Multiple spaces in between like HONG KONG is HONG KONG
6. Port code like SIN-EBGA is Singapore eastbound
7. Misspelling like PYONGTAEK should be PYeONGTAEK
8. abbreviations like SPORE (Singapore); HK is Hong Kong and J.ALI is Jebel Ali

There are other values on destination that non letters or numbers will be discarded.

Initial Results:

Based on the initial results for regression analysis using gradient boosting (GBM) model, draught and ship length are top factors in predicting the travel time (or arrival time). Seasonality (in terms of year and week) is fall below 12% in relative influence factor.

**variable rel.inf**

AVG\_DRAUGHT 26.61848

AVG\_LENGTH 23.93271

AVG\_WIDTH 14.48452

YEAR\_WK 12.02791

AVG\_DIM\_D 11.71585

AVG\_DIM\_C 11.22053

References:

<https://en.wikipedia.org/wiki/Automatic_identification_system>

<http://www.universalcargo.com/shanghais-yangshen-port-suffers-congestion-in-wake-of-carrier-alliances/>