Kpler

Kpler Recruitment Exercises

Kpler Recruitment Exercises

1. The Problem of Ship-To-Ship Cargo Transfer

General context

Most of cargo operations take place between a ship and a land-based terminal. Nevertheless, it sometimes can be useful or necessary to transfer cargo from one ship to another in the open sea and this is called a ship-to-ship (STS) operation.

This operation is the transfer of cargo between seagoing ships positioned alongside each other, either while stationary or underway. One vessel will act as the terminal whilst the other one will moor. The receiving ship is called the daughter vessel and the delivering vessel is called STBL (Ship to be lightered) or Mother vessel. There are some examples about STS:

- animation video,
- video with real vessels.

Practically, these operations usually take place in some specific and designated areas outside of a port, and often happen between a big vessel (for example too large to enter a terminal) and one or more small vessels which transfer the cargo from the big vessel to the terminals, thus the operation is also called `lightering`.

Description of the problem

It is relatively straightforward to track if a vessel has loaded/discharged inside an terminal: we just need to check if it stayed inside the terminal for a sufficient time. A STS is different since they can take place anywhere, between any pair of vessels, so it is harder to track. But they are very important for us since 20-30% of cargo on Crude Oil are transferred via STS: if we ignore them, we cannot model correctly the cargo flows and exchanges.

We want to try to detect these STS using geo-spatial AIS data which allow us to track the cargo vessels. But there are several key issues related to the detection of STS using AIS data:

- The AIS coverage is never perfect, we cannot know exactly where each vessel is at any given time. And different providers may give slightly different information.
- No reliable and complete validation data: no one would tell us if a STS has happened or will happen.
- It is necessary to discover by ourselves which variables are useful and how should we set the parameters given domain knowledge and available data.

Exercise

Instructions

This is kind of open exercise. Given the provided dataset and our description of the STS detection problem, tell us how could you tackle the problem of STS detection. For example, you could tell us:

- What do you think about the provided data set:
  - Which variable seems to be useful or useless to you and why
  - What kind of parameter could we use to describe or discriminate a STS from other shipping operations

- How would you build the model for STS detection:
  - How would you represent a STS (as either a mathematical or programming object)
  - Could you build a prototype algorithm for the STS detection

- Is there any vessel doing STS in the given data set, if yes with which vessel

We are aware of the length of this exercise. We do not expect you to answer all the points above. They are given as source of inspiration and you are free to explore any points you find important or interesting, or even those which you are the most proficient dealing with. If you are really constrained by time, please note that it is more important for us to have a clear description of your reasoning and how you intend to tackle the problem than a hardly readable prototype code which outputs some pairs of STS.

Dataset

Here is a small dataset about the positions over 2 days of 43 vessels, that have passed nearby a sea zone near the port of Galveston, Houston, that is known to have a lot of STS activities and some particularly hard to detect STS cases.

It has already been partially cleaned so:

- There is no input & value error

- There is no obvious incoherence between the coordinates coming from different providers
- If the same position is fetched from several providers then only one is kept, so there is no redundant position

Notes about variables

- `id`: id of the position in our database
- `vessel_id`: identifier of a specific vessel
- `latitude`: geographical latitude of the position, automatically computed by the AIS transmitter
- `longitude`: geographical longitude of the position, automatically computed by the AIS transmitter
- `received_time_utc`: timestamp of the emission of the AIS signal
- `speed`: in knot
- `course`: the cardinal direction along which the vessel is to be steered
- `heading`: the direction that the vessel's nose is pointing
- `draught`: the distance from waterline to keel of a vessel. *This is manually set by the crew*
- `provider_id`: the AIS signal provider's id in our database
- `ais_type`: if the AIS signal come from a radio antenna or a satellite
- `added_at`, `added_by`: when and by which process this data/row was added into our database
- `updated_at`, `updated_by`: when and by which process this data/row was last updated in our database
- `point`: binary that corresponds to a Geometry(Point) of the PostGIS library of PostgreSQL. It uses the coordinate system WGS84 (SRID 4326). It is generated from the given latitude and longitude to facilitate computations.
- `new_navigational_status`: AIS signals include a Navigational Status field which is reported by the vessel. *This is manually set by the crew*. The possible values and their significance are the following:
    - 0 = under way using engine
    - 1 = at anchor
    - 2 = not under command
    - 3 = restricted maneuverability
    - 4 = constrained by her draught
    - 5 = moored
    - 6 = aground
    - 7 = engaged in fishing
    - 8 = under way sailing
    - 9 = reserved for future amendment of navigational status for ships carrying DG, HS, or MP, or IMO hazard or pollutant category C, high-speed craft (HSC)

- 10 = reserved for future amendment of navigational status for ships carrying dangerous goods (DG), harmful substances (HS) or marine pollutants (MP), or IMO hazard or pollutant category A, wing in ground (WIG)
- 11 = power-driven vessel towing astern (regional use)
- 12 = power-driven vessel pushing ahead or towing alongside (regional use)
- 13 = reserved for future use
- 14 = AIS-SART (active), MOB-AIS, EPIRB-AIS
- 15 = undefined = default (also used by AIS-SART, MOB-AIS and EPIRB-AIS under test)

Q&A for frequently asked questions

- Could STS happens while the vessels are anchored or while navigating?

Both are possible. We have seen vessels doing STS at full stop or while sailing.

- What is the duration of a STS?

It depends, in practice, a STS could ranges from 1-2 hours up to a day.

- What is the spatial precision for the AIS data?

The average precision for the coordinates is around 0.0001 arc minutes. It is negligible comparing to the size of the vessels for our problem.

- Why is there such a disparity between the position coverage between the vessels?

We had not deliberately obfuscated the data to make it difficult. Sometimes we just lost track of the vessels: maybe there was a system failure, or maybe the captain decided to switched it off for… *diverse reasons*. Trying to find what happened when there is no direct data is one of the major problems we are dealing with.

2. Oil Product Storage Pipeline Architecture

Kpler wants to detect storage variations over time of oil products in petrol tanks. To build those timeseries per storage site we use satellite images to monitor petrol tanks with moving roofs. We use the shadows to calculate the volume in each tank. Kpler has signed contracts with satellite image providers, images are available via APIs. We would like to follow 10 sites first, then 300 in the long run. The frequency varies per site, it coult be 1 image every few days, or few weeks. The user result is a dashboard where user can select a site, or a broader zone (containing several sites) and see the resulting linechart, with volume level for days we have an image, and extrapolated data the other days.

Describe a **pipeline architecture** (text or schema) and the technologies/tools you would use for this project.

Few comments to guide your architecture:

- Explain which raw data and results will you store. Which datastore will you use? What data model?
- Sequential/Parallel processing?
- Where do you see bottleneck or potential issues?

Note: We don’t expect you describe the algorithms to extract the information from the images.

To give you an idea, here is an oil tanks satellite picture.

3. Expected Output

Please email us :

- Your question answers
- Your code
- All other important files

Please do not upload your result on a public repository or website.

Good luck :)

This project is maintained by Kpler