



Learning from our movements – the Mobility Data Analytics era

Yannis Theodoridis

Data Science Lab., Univ. Piraeus

MASTER Workshop @ECML/PKDD Conf.,
Würzburg, Germany, 2019.09.16



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Outline

1. About movement data

2. A flashback to the past (from a personal perspective)

3. Mobility data analytics pipeline

- Getting familiar with the data
- Data pre-processing (cleansing, transformation, enrichment, etc.)
- Discovering hidden knowledge (group behavior analytics, predictive analytics, etc.)

4. A real-world use case

- ... from the aviation domain

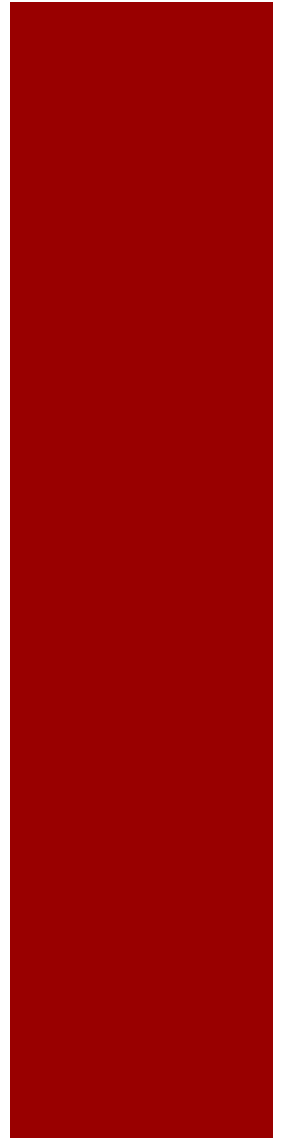


1. About movement data

2. A flashback to the past

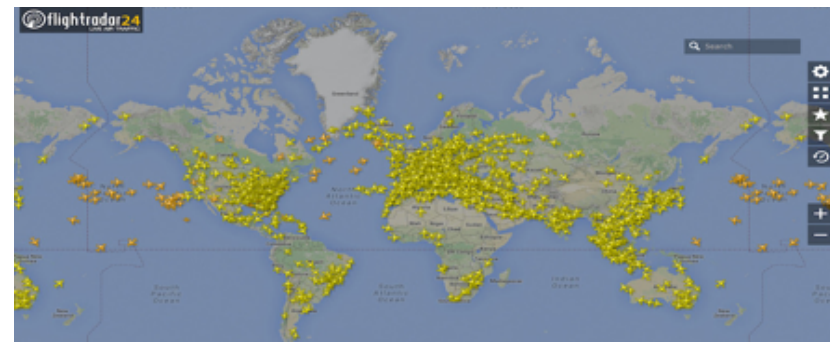
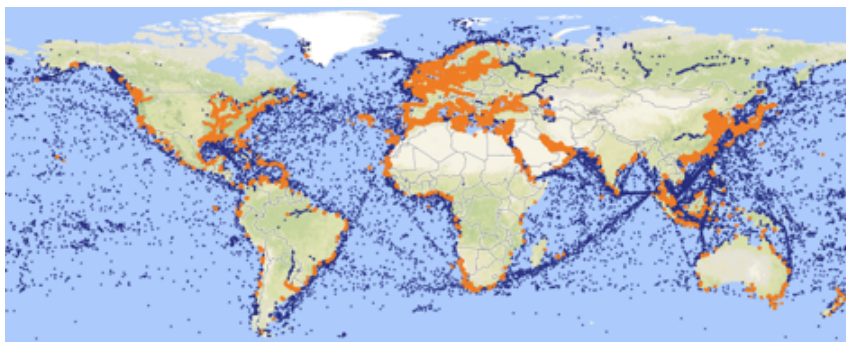
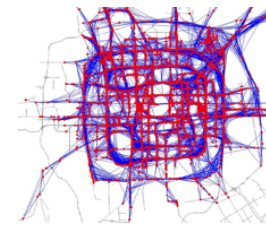
3. Mobility data analytics pipeline

4. A real-world use case



Some questions, to start with...

- Can we find **available** movement datasets out there?
- How much of this data is **useful** for analytic purposes?
- What kind of analytics is **suitable** for this data?
- Are these datasets **large/complex** enough, for us to call them “Big Data”?



Examples of datasets @ urban domain

- **NYC taxis** (source: NYC Taxi & Limousine Commission): 1.4 billion trips, Jan. 09 – Dec.17.
- **Ride-hailing app** data are also provided
- Attention: only pickup – drop-off locations are available

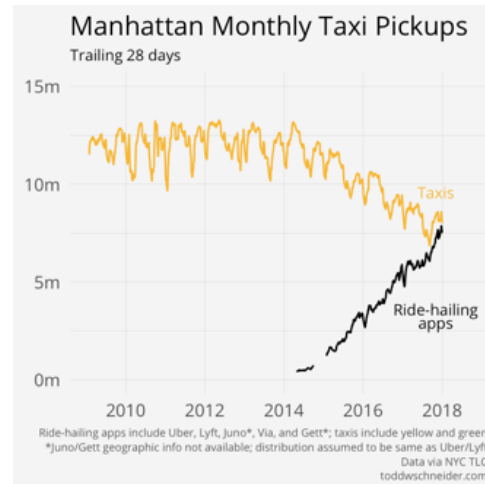
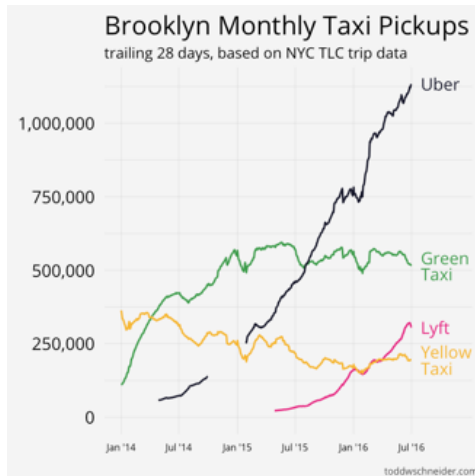
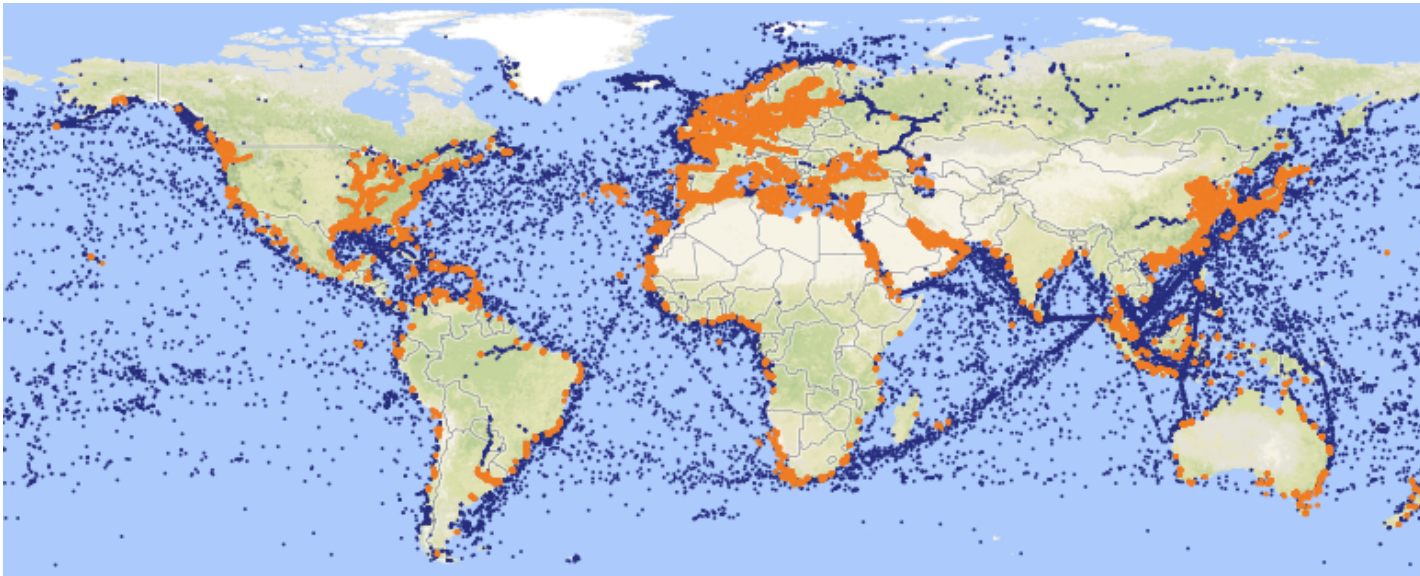


image source: toddwschneider.com

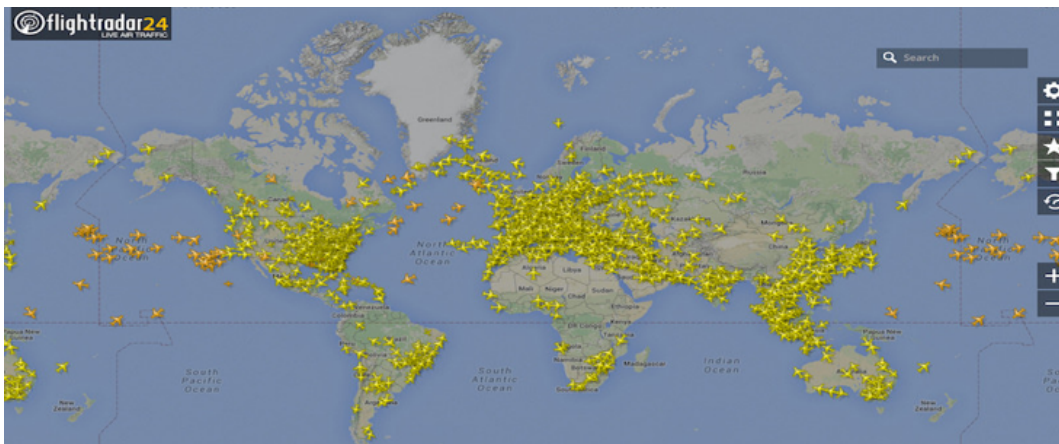
Examples of datasets @ maritime domain

- **AIS** (Automatic Identification System) **tracking system** for identifying & locating vessels at sea
 - 400,000 vessels worldwide (source: vesseltracker.com)



Examples of datasets @ aviation domain

- **ADS-B** (Automatic Detection System-Broadcast) **tracking system** for identifying & locating aircrafts on air
 - 15,000 aircrafts (average) fly at the same time worldwide (source: flightradar24.com)



“Aireon is providing the first global air traffic surveillance system using a satellite-based, space-based Automatic Dependent Surveillance-Broadcast (ADS-B) network ... The data is broadcast every half a second from the aircraft, and is being used by Air Traffic Controllers (ATCs) to identify and separate aircraft in real time.” source: [aireon.com](https://www.aireon.com).

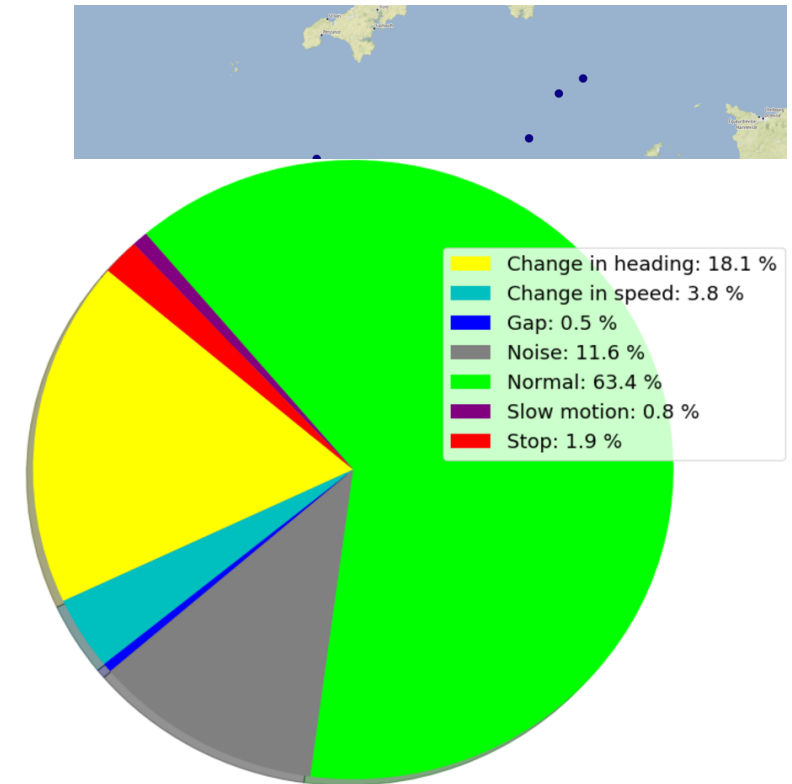
From raw data to useful information

An example: AIS maritime dataset (Ray et al. 2018)

- 6 months in Brest area, ~19M AIS records
- Collected by Naval Academy, Brest (FR)

Findings (Theodoropoulos et al. 2019):

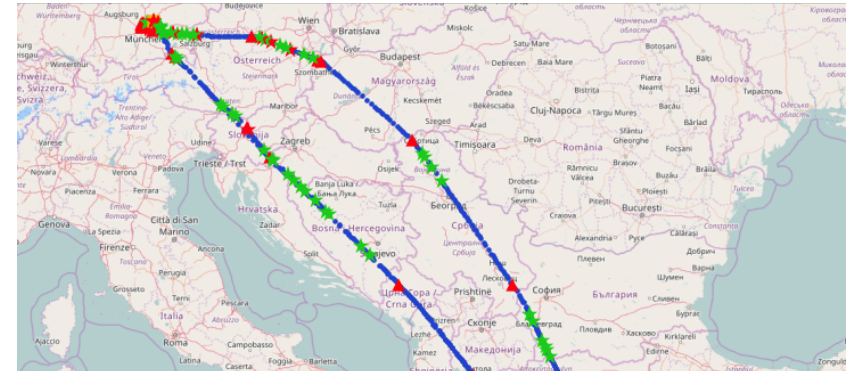
- Only 1/3 of records correspond to locations outside ports
- Out of them, only 1/4 appear to be useful for analytic purposes
- The tentative useful ones can be enriched to be made really useful for analytics
 - e.g. annotated according to their status w.r.t ports



From raw data to useful information (cont.)

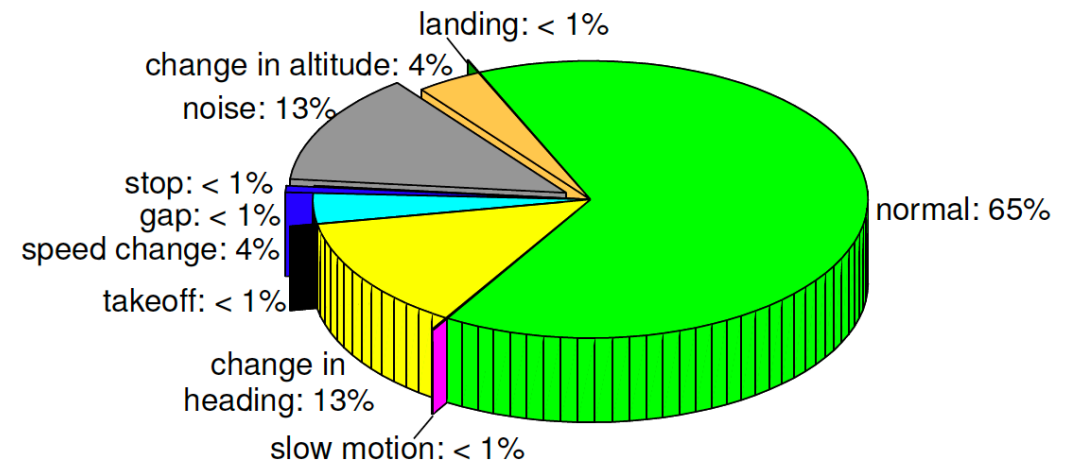
Another example: ADS-B and IFS-radar aviation datasets

- ADS-B (flightaware.com; www.adsbhub.org): aircraft positions over Europe during 1 day
- IFS-radar (www.enaire.es): aircraft positions over Spain during 1 week



Findings (Patroumpas et al. 2018):

- Only 1/3 of records correspond to 'interesting' behavior, hence useful for analytic purposes
- Can be annotated by e.g.
 - their motion status, their proximity to airports, etc.



1. About movement data

2. A flashback to the past

3. Mobility data analytics pipeline

4. A real-world use case

