Semantic Aware Anomaly Detection in Real World Parking Data

Arnamoy Bhattacharyya[1], Weihan Wang[2], Christine Tsang[2], Cristiana Amza[1]

[1] University of Toronto, [2] Smarking Inc

Motivation

PRESS RELEASES

Drivers Can Find Parking Faster with New INRIX On-Street Parking

- *First integrated service to help drivers find available street parking in major cities worldwide*
- *Service also helps cities better monitor and manage street parking inventory to improve urban mobility*
- *Expands market leading Off-Street Parking service that provides insight into pricing and availability for more than 80,000 parking lots and garages in 31 countries*

Novi, Mich. – Telematics Automotive 2015 – June 3, 2015 – Everyone who has ever been

Parkloco Is Transforming Parking Industry through Data Analytics

New parking data analytics startup announces its seed funding.

BOSTON, MA (PRWEB) APRIL 14, 2016

Across the United States, up to 30% of downtown congestion is caused by vehicles searching for parking spaces. This issue is not necessarily due to the lack of parking spaces but rather, inefficient data within the parking ecosystem. Parking facility owners and managers have tremendous opportunities to optimize their facilities' performance through real-time data made available by Parkloco's enterprise software.

Unlike most companies that are attacking the parking problem from the consumer end, Parkloco aims to optimize parking facility occupancy for operators, maximize revenue capture and performance for owners, and improve the overall parking experience for both parties.

The Brown University spin off parking startup based in Boston, today announced that they have raised an undisclosed Seed round of

Smart parking, or Smarking, making its mark in NYC

BY **REW** • JULY 7, 2016

City parking took a technological leap last week when two industry giants formed a new partnership.

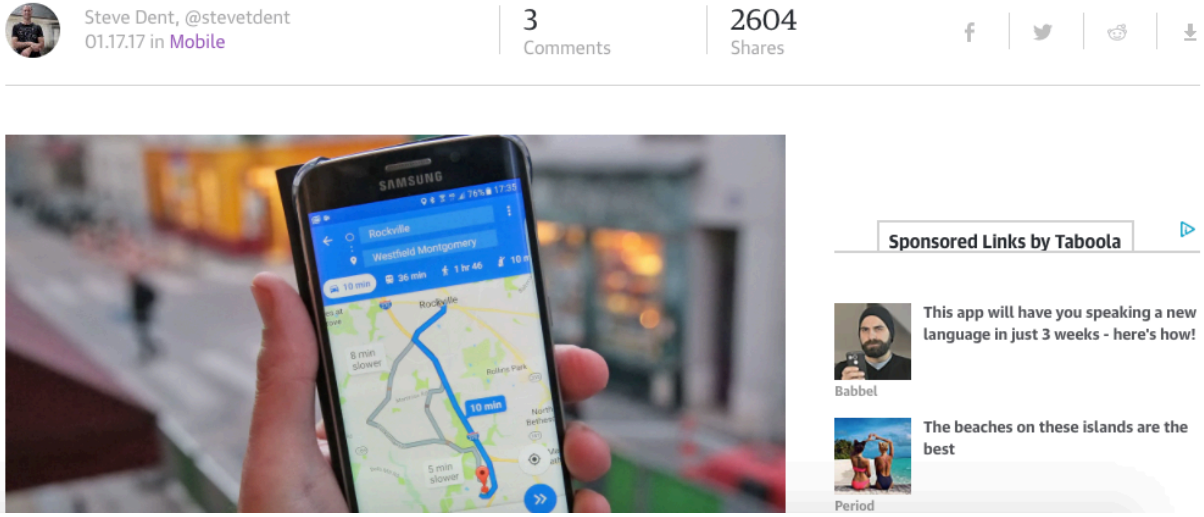
Smarking, a provider of data analytics software to the parking industry, and Citizens Parking's Icon division, New York City's largest operator of parking facilities, will harness the power of

REAL ESTATE WEEKLY

Motivation

Google Maps may soon offer parking suggestions

Knowing where spots are available will make route planning easier.



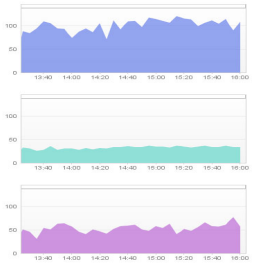
https://www.engadget.com/2017/01/17/google-maps-may-soon-offer-parking-suggestions/

Architecture of Data Analytics Systems

Architecture of Data Analytics Systems

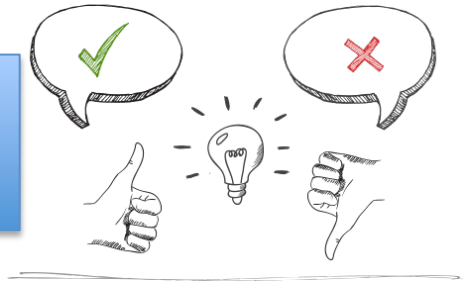
Sensors

Monitored data at Clients

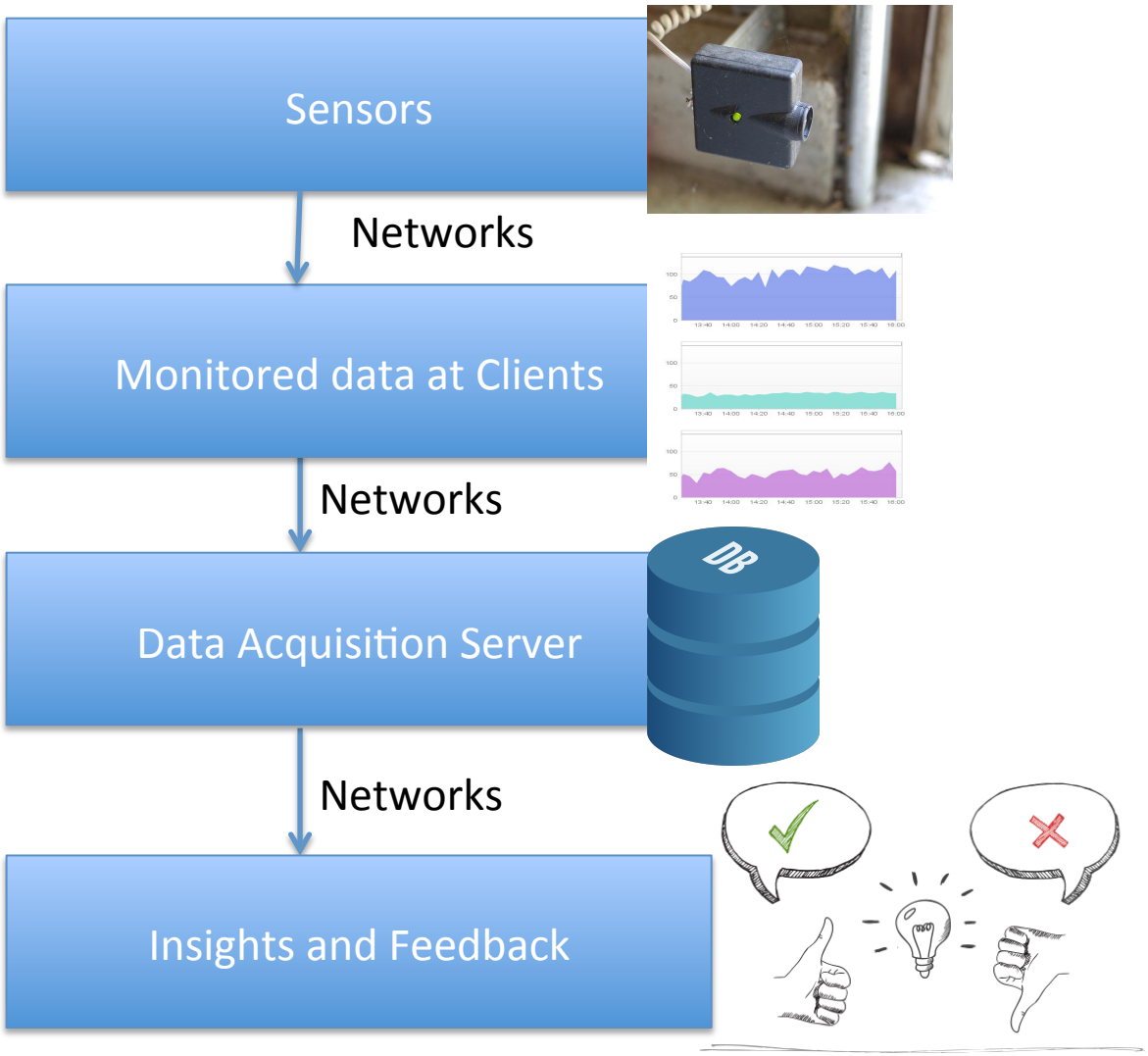


Data Acquisition Server

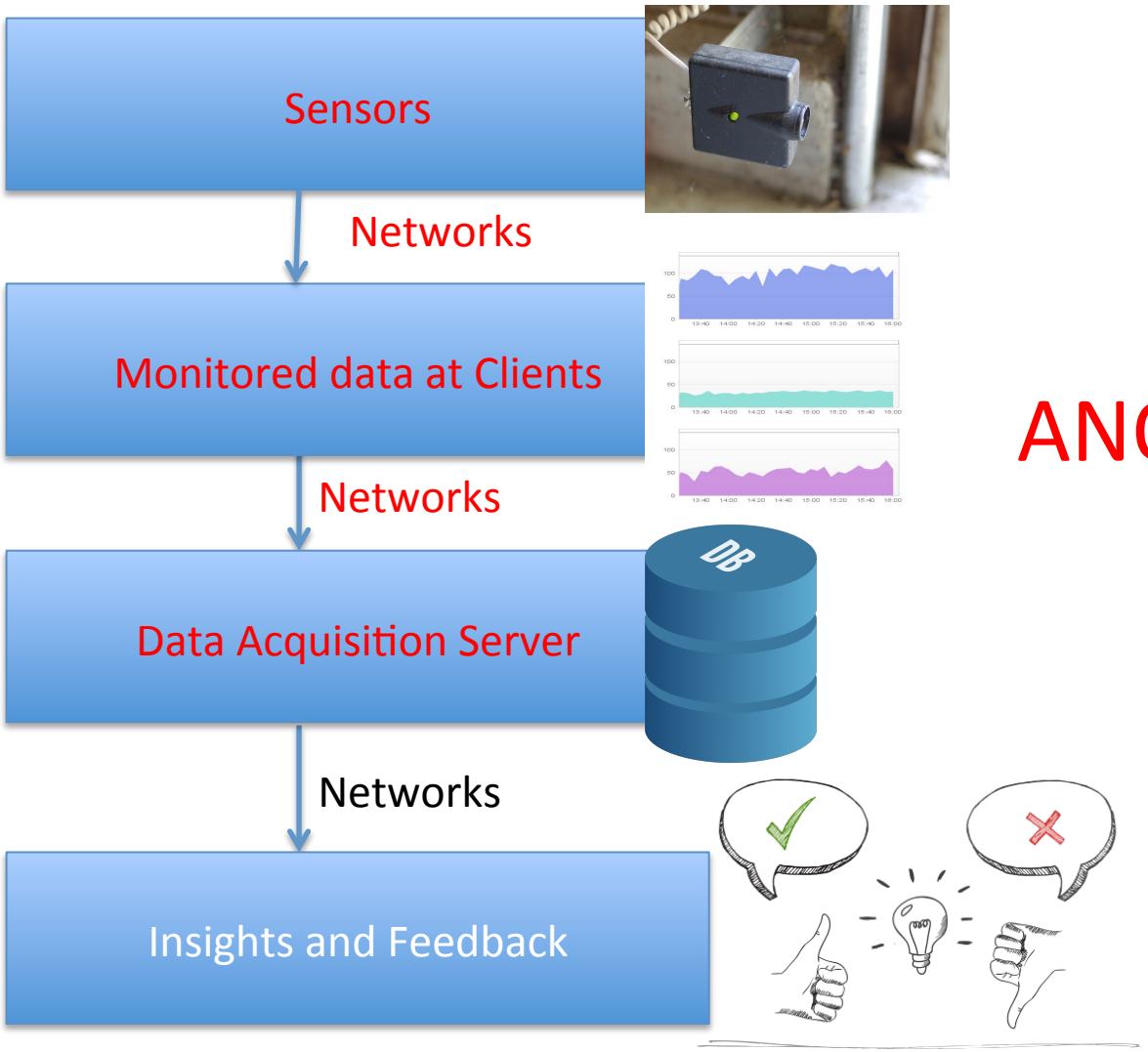
Insights and Feedback



Architecture of Data Analytics Systems

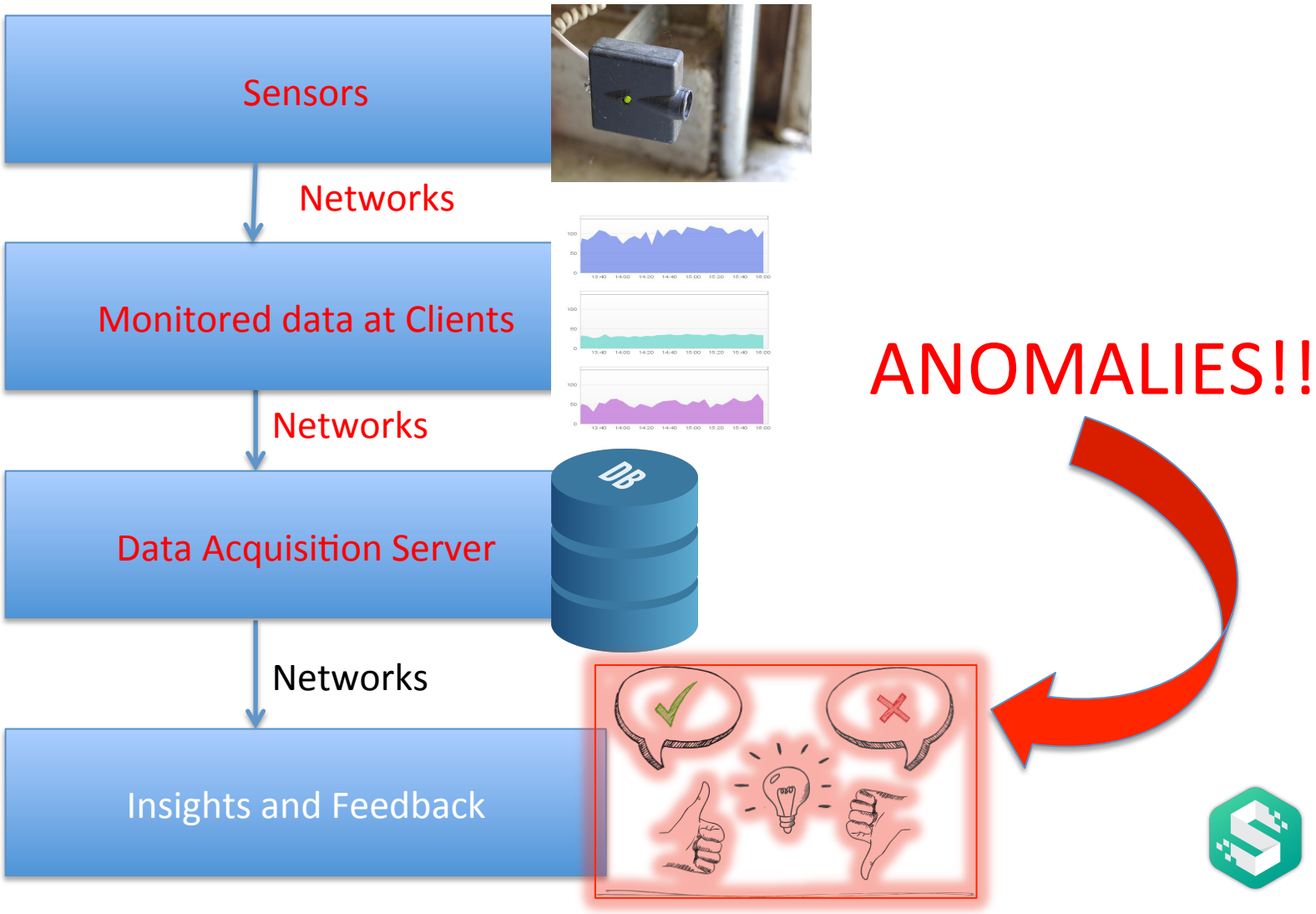


Architecture of Data Analytics Systems

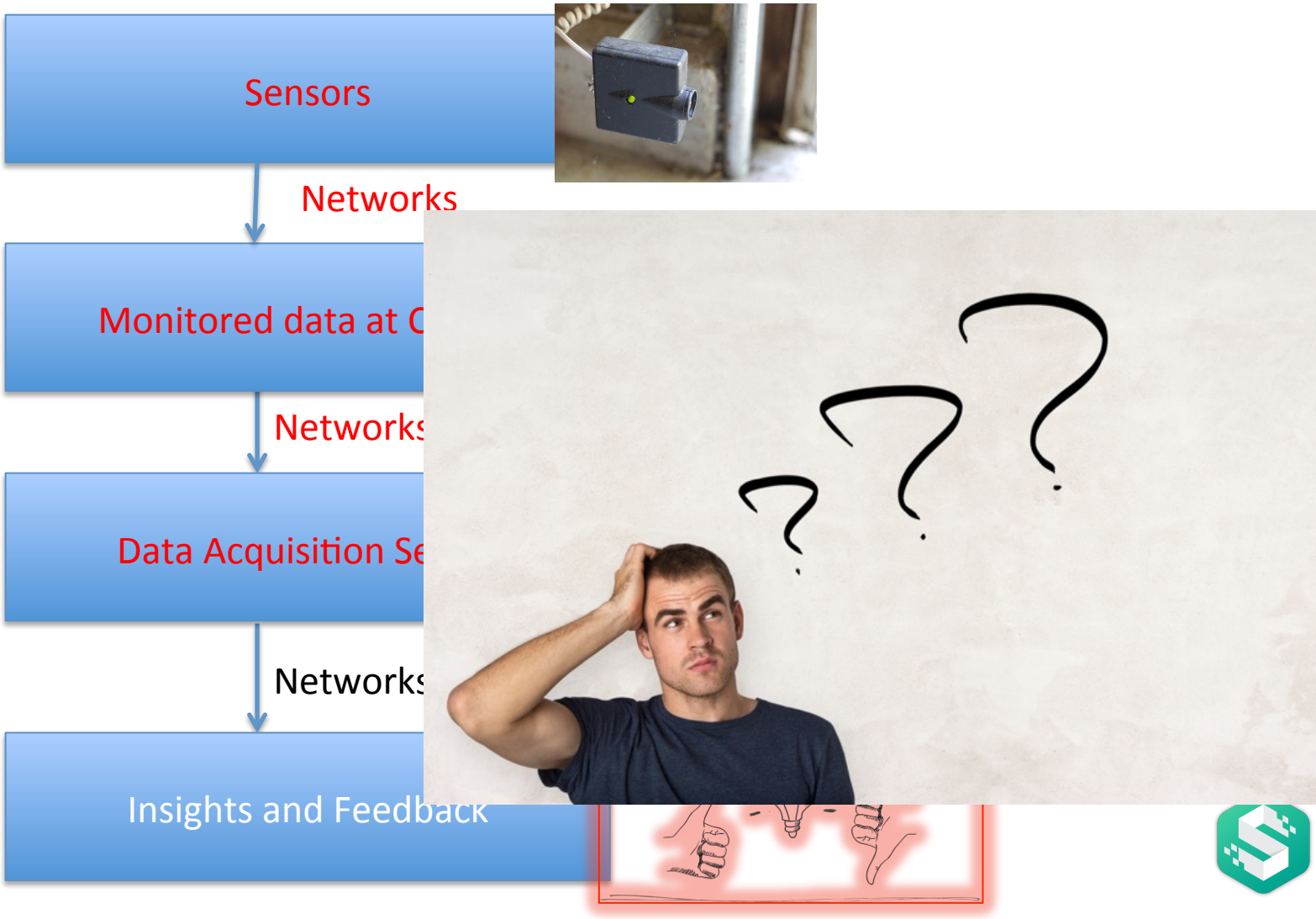


ANOMALIES!!

Architecture of Data Analytics Systems



Architecture of Data Analytics Systems



Architecture of Data Analytics Systems

Fast detection of anomaly is necessary

Fast and Real-time identification of Anomalies

Technique overview

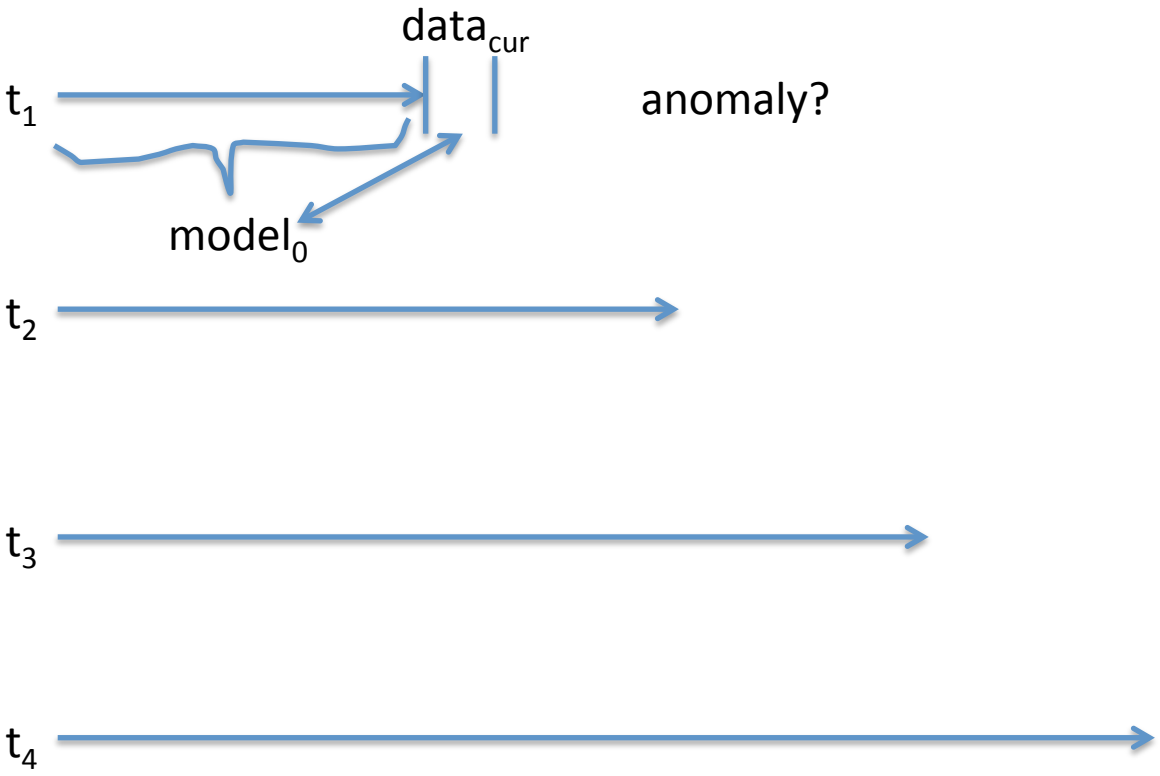
Three step process:

- Learn pattern from history
- Compare with current data
- If different → Anomaly (raise alarm)

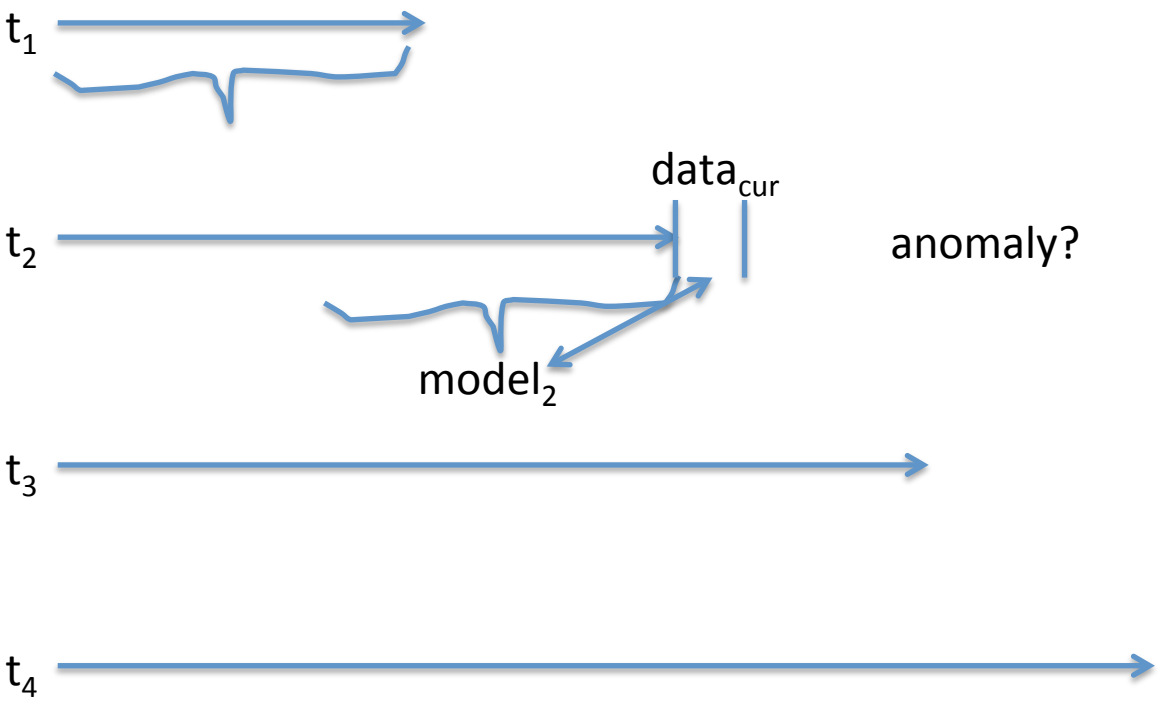
Challenge: Growing History

- Real time data, continuously arriving
- Size of history changes
- Learn from a *sliding window* of history

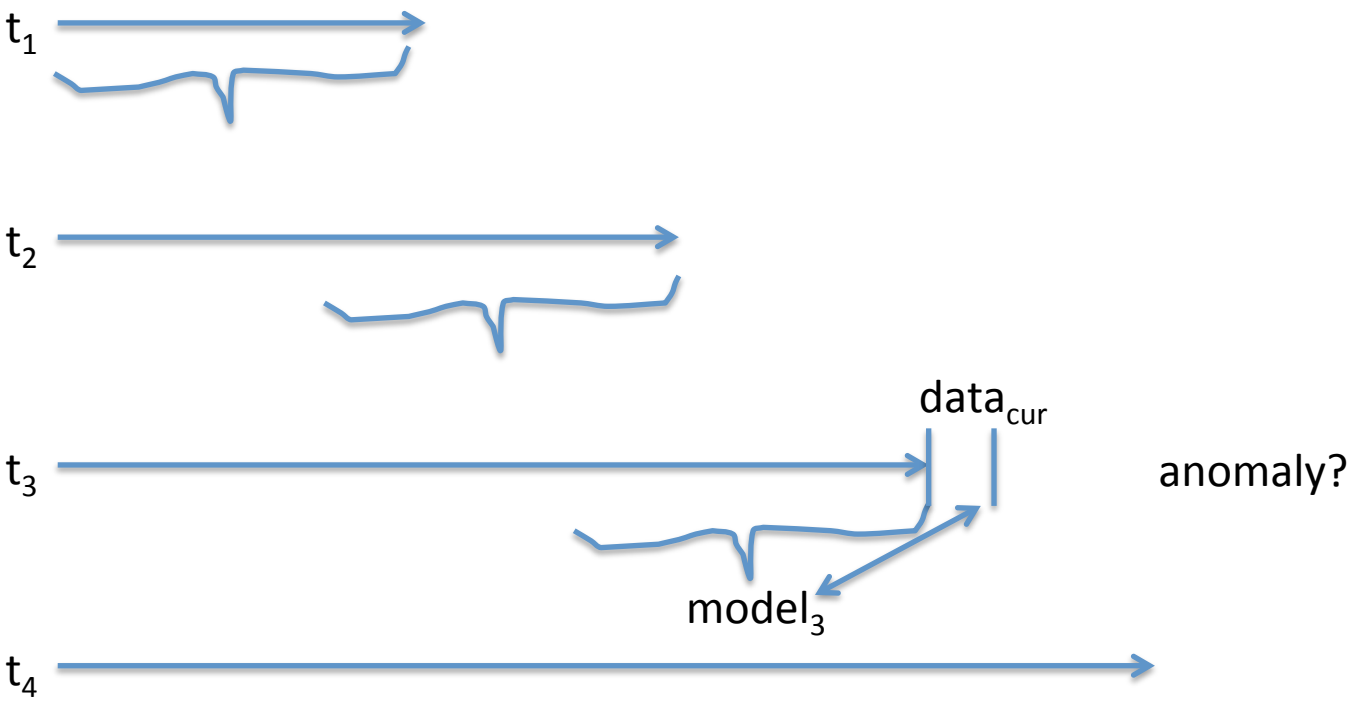
Sliding Window Based Anomaly Detection



Sliding Window Based Anomaly Detection



Sliding Window Based Anomaly Detection



Challenge

- Lot of re-learning (at regular intervals)
- Necessary to be **fast** for real time analytics
- We derive techniques for faster learning

Fast Learning from Historical Data

- Goal: Learn normal behaviour from history *fast* and compare with current

- Observation: Historical data is not clean (anomalies within)

- Method:

→ Clean up historical data (remove anomalies)

→ Learn normal pattern from ***reduced*** dataset.

Learning from Historical Data

How to perform data-cleaning?

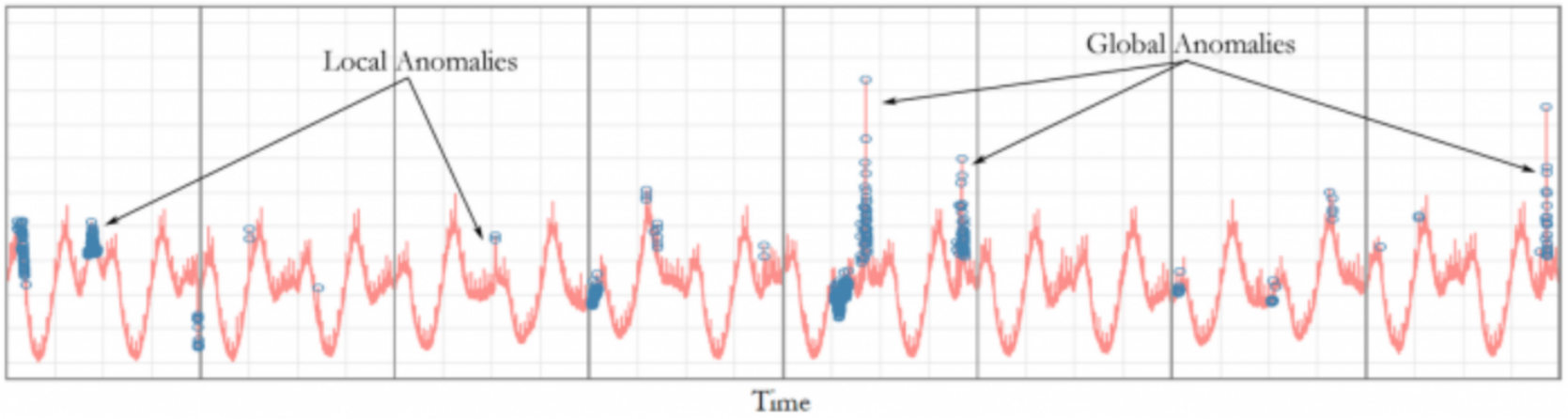
Two main types anomalies*:

1. Global Anomalies
2. Local Anomalies

* https://blog.twitter.com/engineering/en_us/a/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series.html

Learning from Historical Data

Global and Local Anomalies



Learning from Historical Data

- Global Anomalies: Simpler (faster method to detect)
- Local Anomalies: Complex (slower method to detect)

Learning from Historical Data

- Global Anomalies: Simpler (faster method to detect)
- Local Anomalies: Complex (slower method to detect)
- Method:

Step 1: Detect global anomalies (using IQR)

Step 2: Remove them from dataset

Step 3: Detect local anomalies from a reduced dataset (Signal Processing + Machine Learning)

Global Anomaly Identification

- Stand out from dataset
- Some aggregation on the dataset and then use IQR

- Aggregation
  - Data collected every 15 minutes
  - Get/aggregate weekly average

Global Anomaly Identification

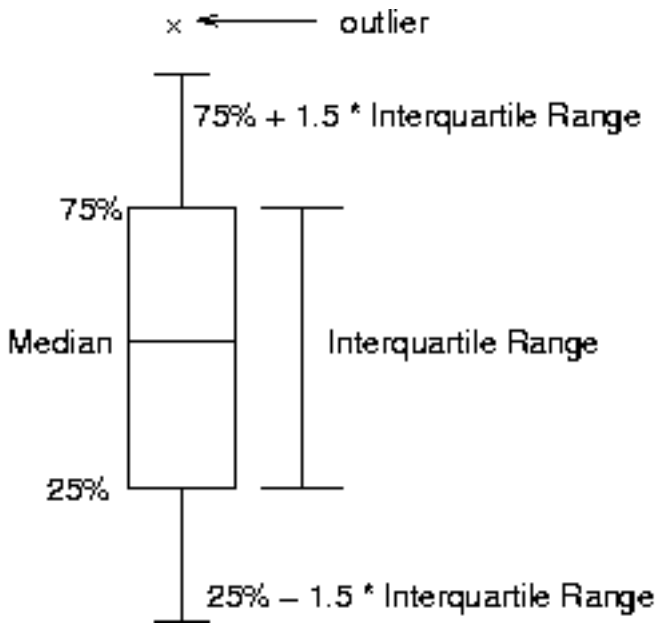
- Stand out from dataset
- Some aggregation on the dataset and then use IQR

- Aggregation
  - Data collected every 15 minutes
  - Get weekly average
    - e.g. For a data of 10 year history, reduce the dataset size from 350K to 520

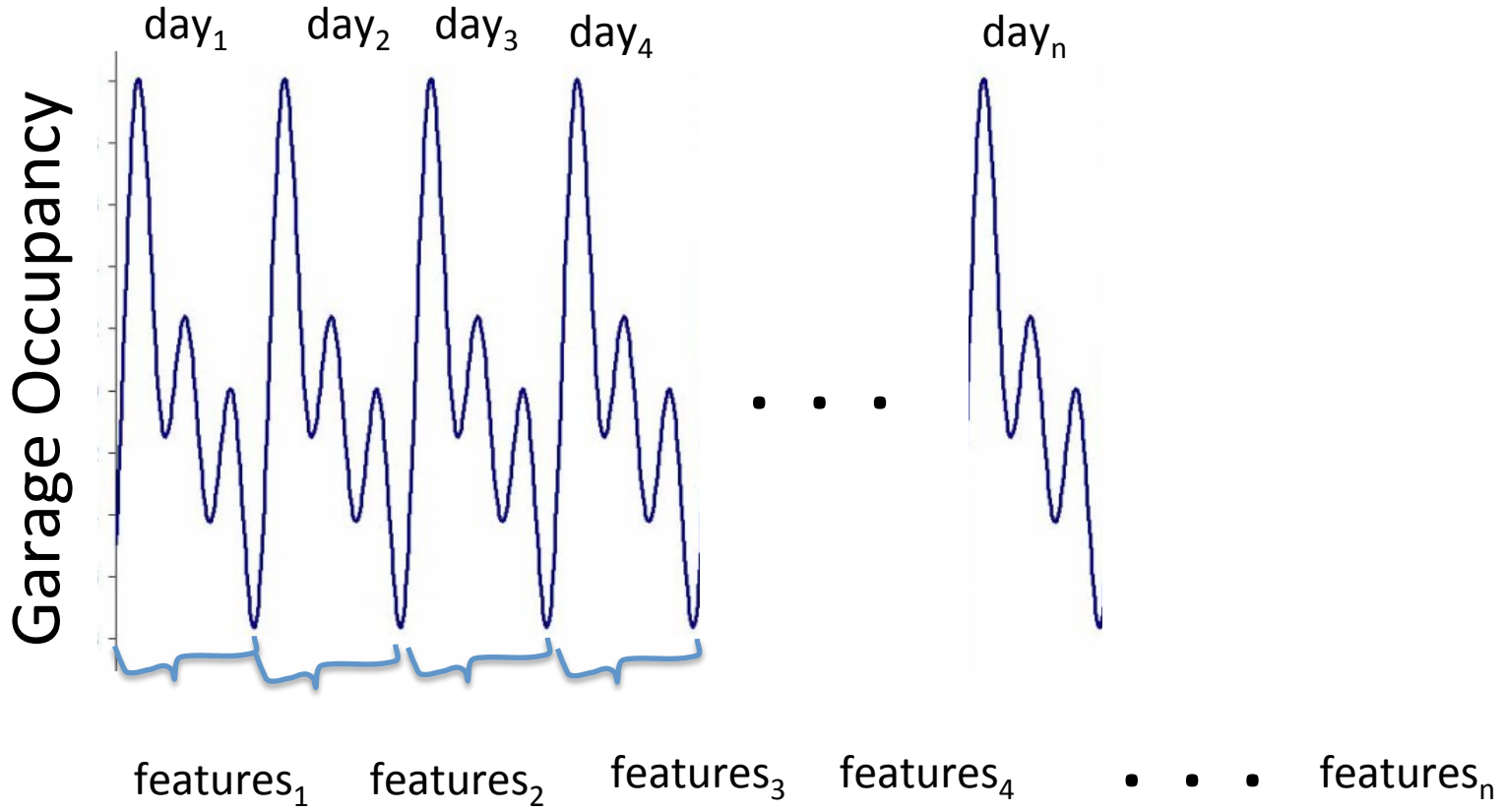
Global Anomaly Identification

- Stand out from dataset
- Some aggregation on the dataset and then use IQR

- IQR
  - Standard statistical outlier detection method



Local Anomaly Identification



Real Time Anomaly Detection

day_{test}

$features_{test}$

SVM Classifier

Normal

Anomalous

Experimental Evaluation

- Santa Monica (California, USA) open source parking data*
- Real time data collected for 12 garages, every 5 minutes
- Historical Data from Nov 6th 2014 till Jan 23rd, 2017
- Real Time Anomaly detection for Jan 24th – Jan 31st 2017

* https://data.smgov.net/Transportation/Parking-Lot-Counts/ng8m-khuz/data

Results

- Quite a few anomalies in historical data (Total 1894 anomalous days)
- Data cleaning sped up the learning by upto 3.1x
- Detected 32 real time anomalous days (verified by looking at the signals manually).

Results

- Does it work for other datasets?

San Francisco, CA
Startup by a group of MIT graduates

No false negatives: All anomalies previously manually detected by humans were identified.

False positive number was also low.

Conclusions

- Presented an approach for real time anomaly detection on parking data
- Based of global and local anomaly detection.
- Works for multiple real-world dataset.

Thank you. Questions?

- http://www.eecg.toronto.edu/~arnamoyb/
- arnamoyb@ece.utoronto.ca