SP Orchestration & Automation: Connected Car Analytics with Machine Learning

Dr. Mehdi Nikkhah, Research Software Engineer – Data Scientist, Innovations Lab - CTAO

Patricia Sampedro Garcia, Software Engineer, CTAO

Chris Metz, Distinguished Engineer, CTAO

April 5, 2017

Additional Team Members

- Herb Wildfeuer, Principle Engineer, Innovations Lab - CTAO
- Tyler Levine, Sr. Software Engineer, CTAO
- Quentin Chen, Software Engineer, CTAO
- Daniel Malachovsky, Sr. Software Engineer, CTAO
- Stanislav Jamrich, Software Engineer, CTAO
- Lubomir Balogh, Software Engineer, CTAO
- Andrej Vanko, Project Manager, CTAO

Problem Statement (1)

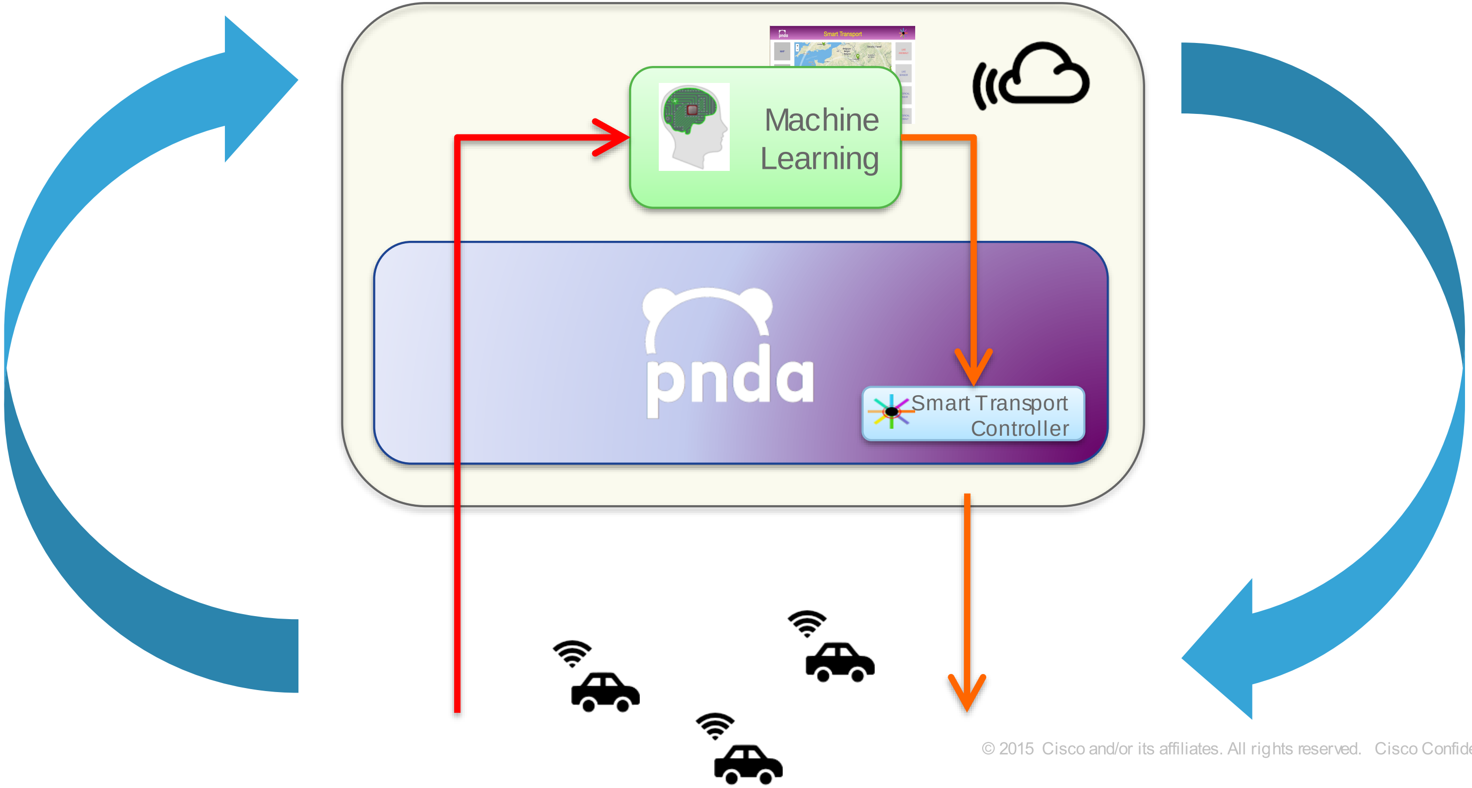
- Connected Cars will generate 25GB of Sensor Data per Day per Car for cloud ingestion and processing
- WAIT … WHAT?!?
- Automakers/OEMs, drivers, smart cars, etc. want to avoid “problems”
- Problems → Higher Costs → Unhappiness

Problem Statement (2)

- All of this Car **BIG DATA** in the cloud might have **SMALL DATA** telling us about current and future problems
- Avoid Problems → Lower Costs → Happiness
- So what is needed here to find the small data pointing to current and future problems

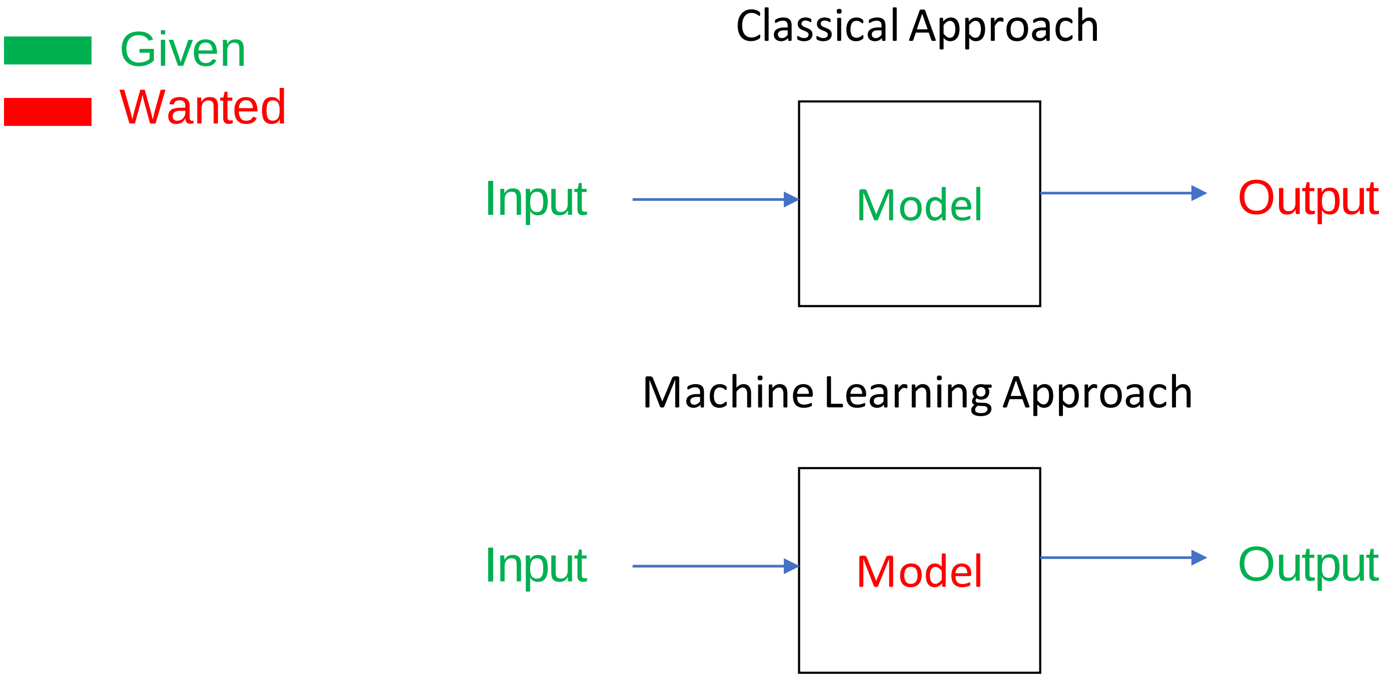
Machine Learning!!

CC Analytics Cloud, Machine Learning and the "Virtuous Circle"



Machine Learning

Machine Learning vs. Classical Approach



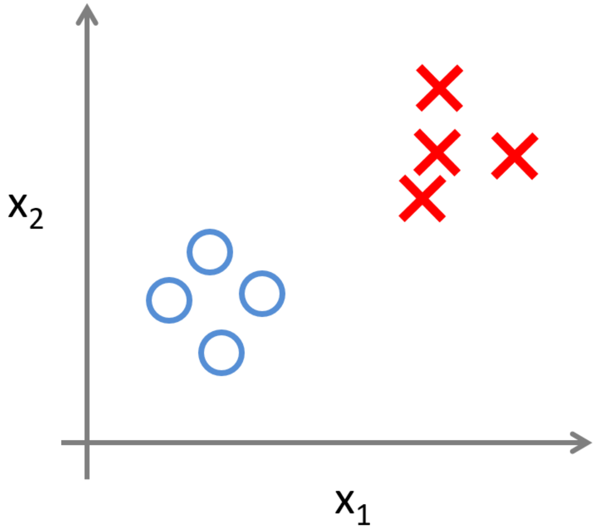
How Machine Learning Works

- Machine Learning builds a model from the data
  - Supervised: Data and Labels
  - Unsupervised: Data with no label

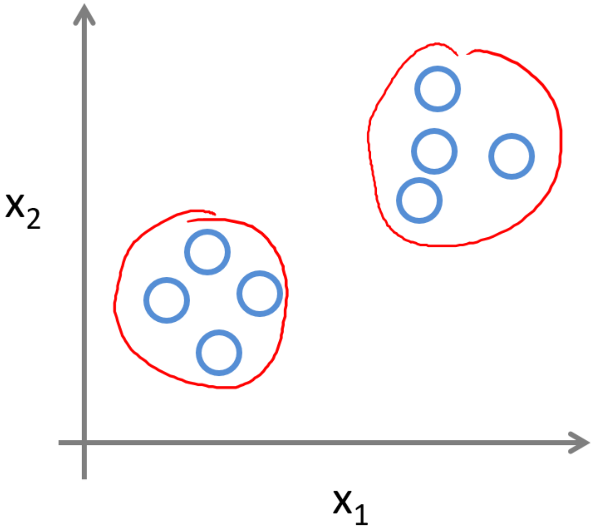
- The model is used then to:
  - Predict the outcome of a system
  - Recognize complicated patterns in the new data points
  - Classify inputs

Supervised vs Unsupervised Learning

Supervised Learning

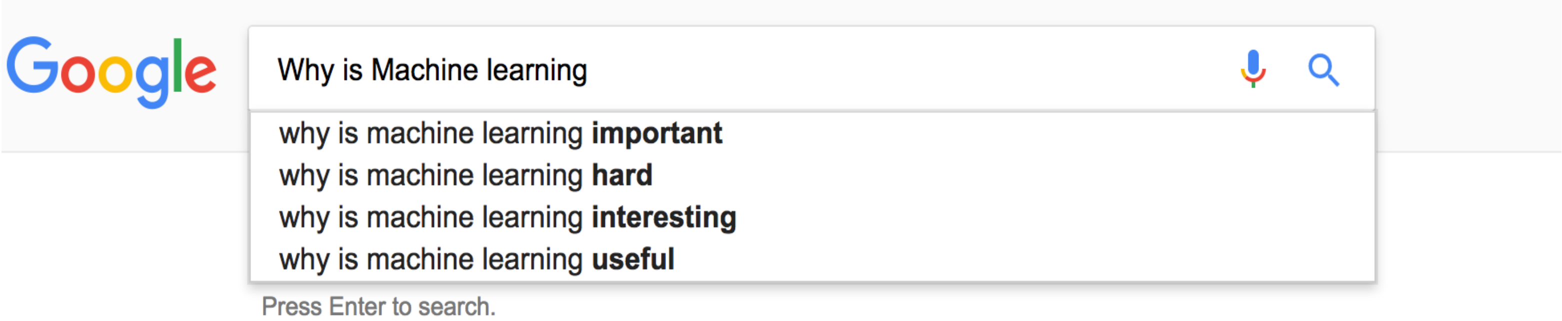


Unsupervised Learning

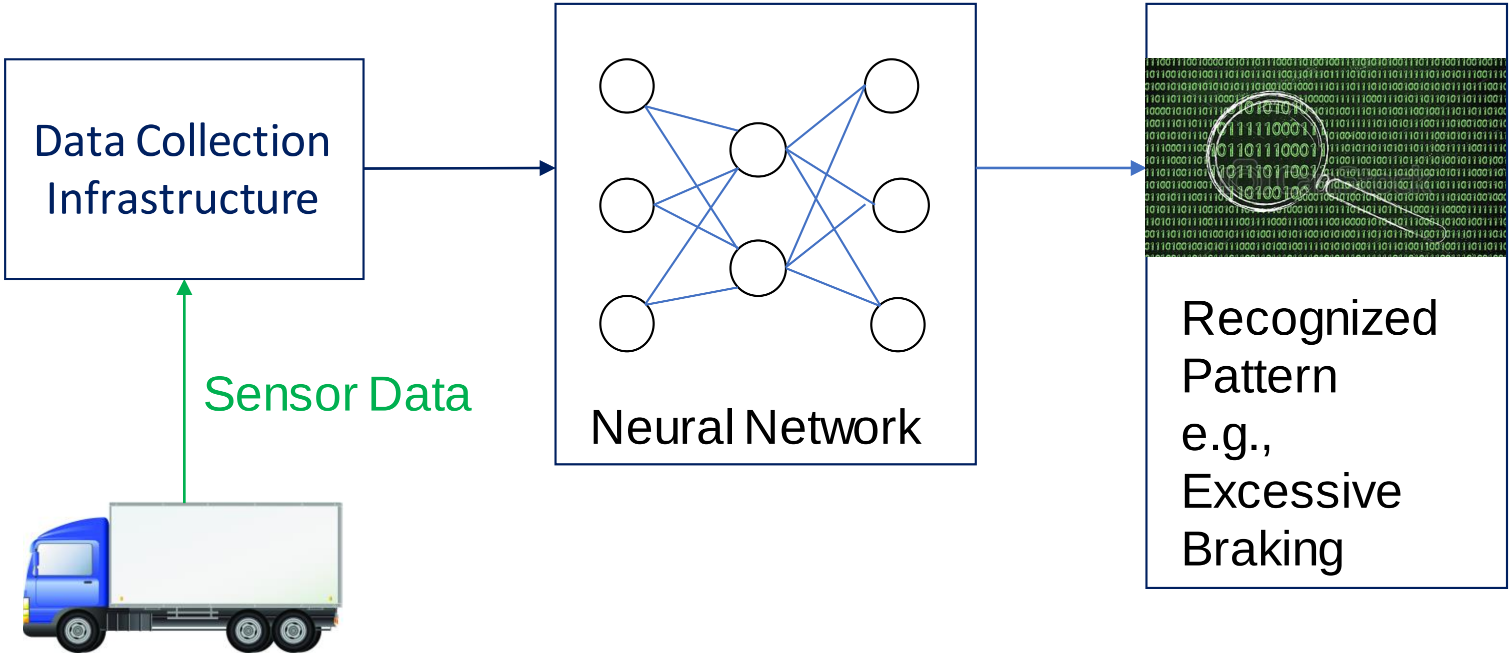


Supervised Learning

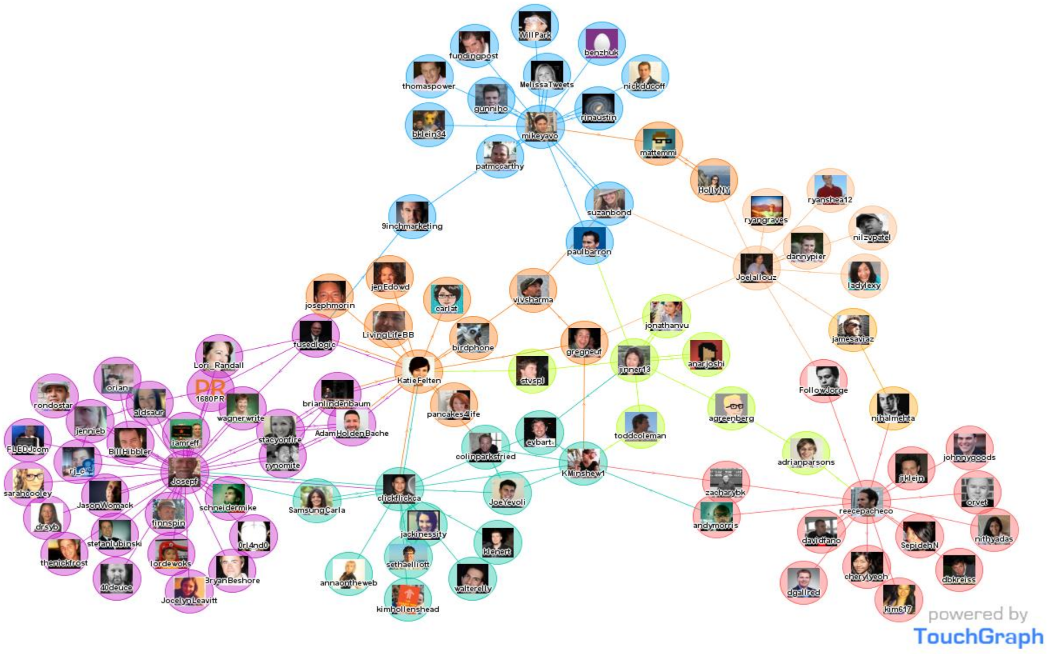
Example: Neural Network for Prediction



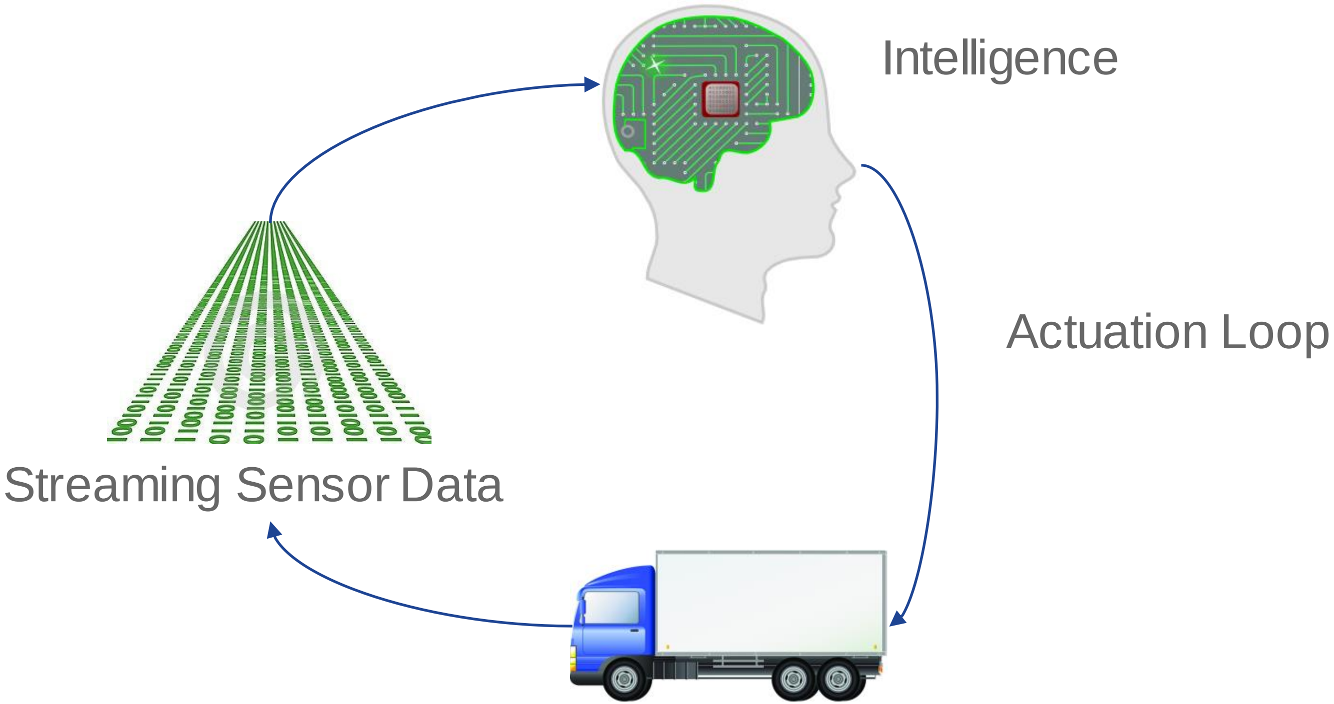
Example:
Neural Network for Pattern Recognition



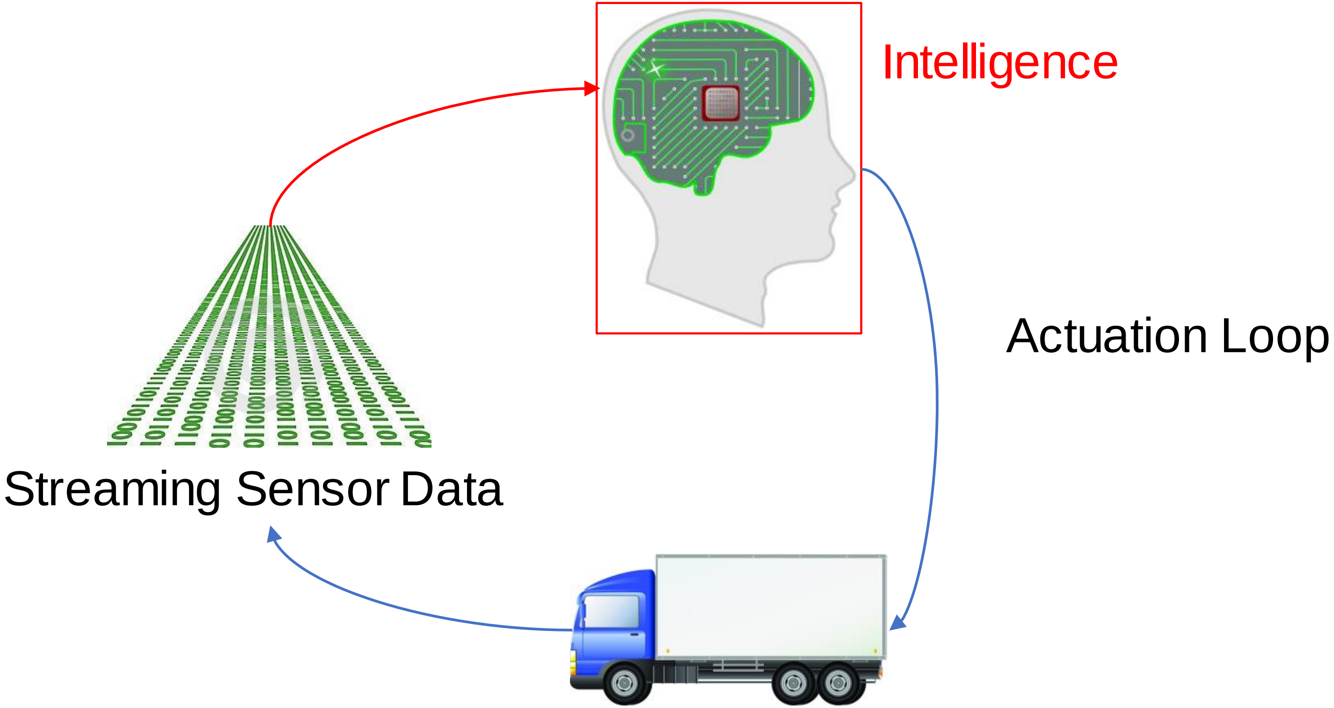
Example: Clustering (Unsupervised)



Automation Use Case: Connected Car



Automation Use Case: Connected Car



Do we always need ML?

- When not to use ML?
  - If the problem is easily solved with simple algorithms
  - *e.g.*, Thresholding, regression, etc. work better

- When to use ML?
  - If the problem is non-linear, features that affect the outcome are not completely known, complicated patterns, data is ample

Rule-based Analytics

- Driving in dangerous conditions is a complex problem
  - There is no strict definition of "unsafe"
    - *e.g.*, some consider driving 55 mph on a rainy day dangerous, some don't!
  - However, there is consensus about driving "fast" on a rainy day being dangerous
  - The variability lies in the definition of "fast"

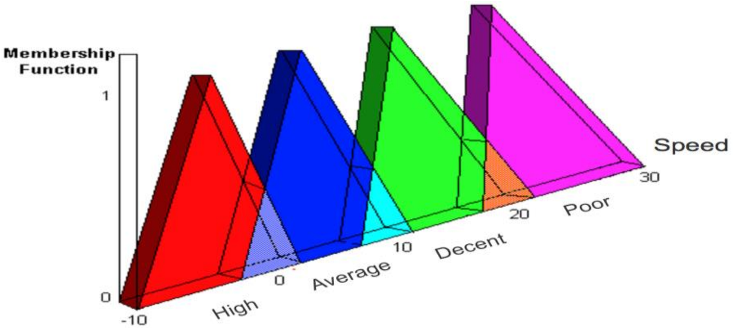
Fuzzy Logic

- Fuzzy Logic variables:
  - Speed of the car (high, average, decent, poor)
  - Current conditions of the road (high, average, decent, poor)
  - Current weather (sunny, cloudy, foggy, rainy, snowy)

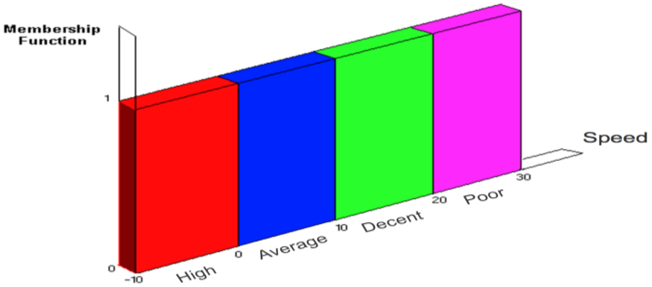
- Examples:
  - **If** speed is HIGH **and** conditions of the road are POOR **and** the weather is SUNNY **then** UNSAFE.
  - **If** speed is AVERAGE **and** conditions of the road are GOOD **and** the weather is FOGGY **then** SAFE.

Fuzzy Logic contd.

- Fuzzy Logic goal: to model logic reasoning with imprecise statements
  - e.g., ‘speed is High’, is a vague sentence that is determined by the true value of the variable speed



Variables on Fuzzy Logic

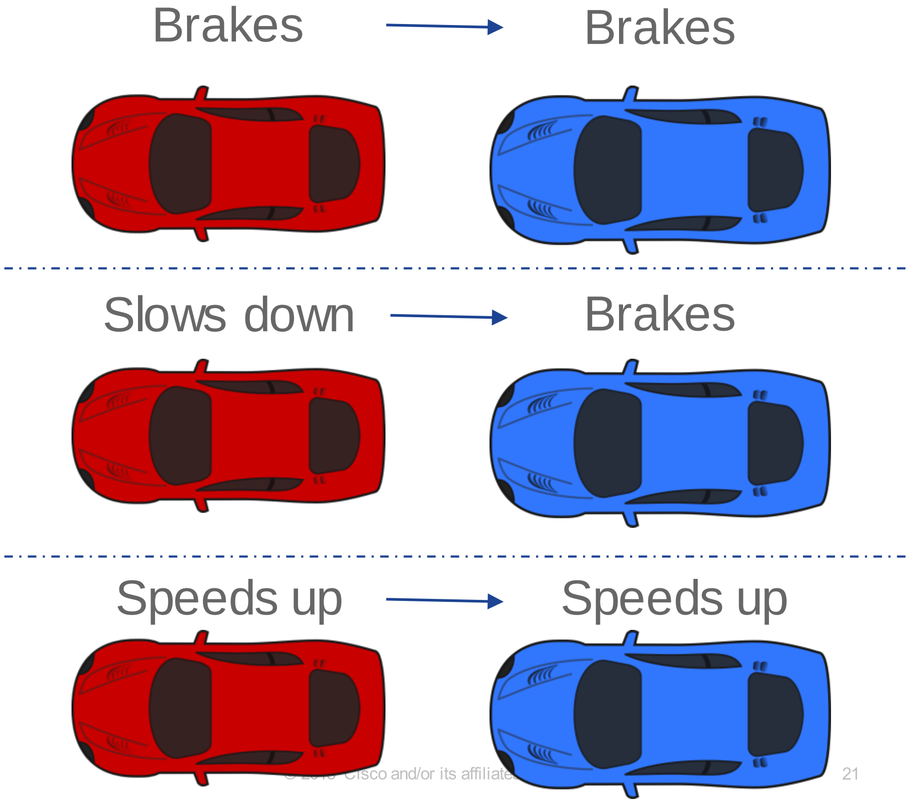


Variables on Boolean Logic

Machine Learning and Connected Cars Discussion

Complicated Pattern Recognition

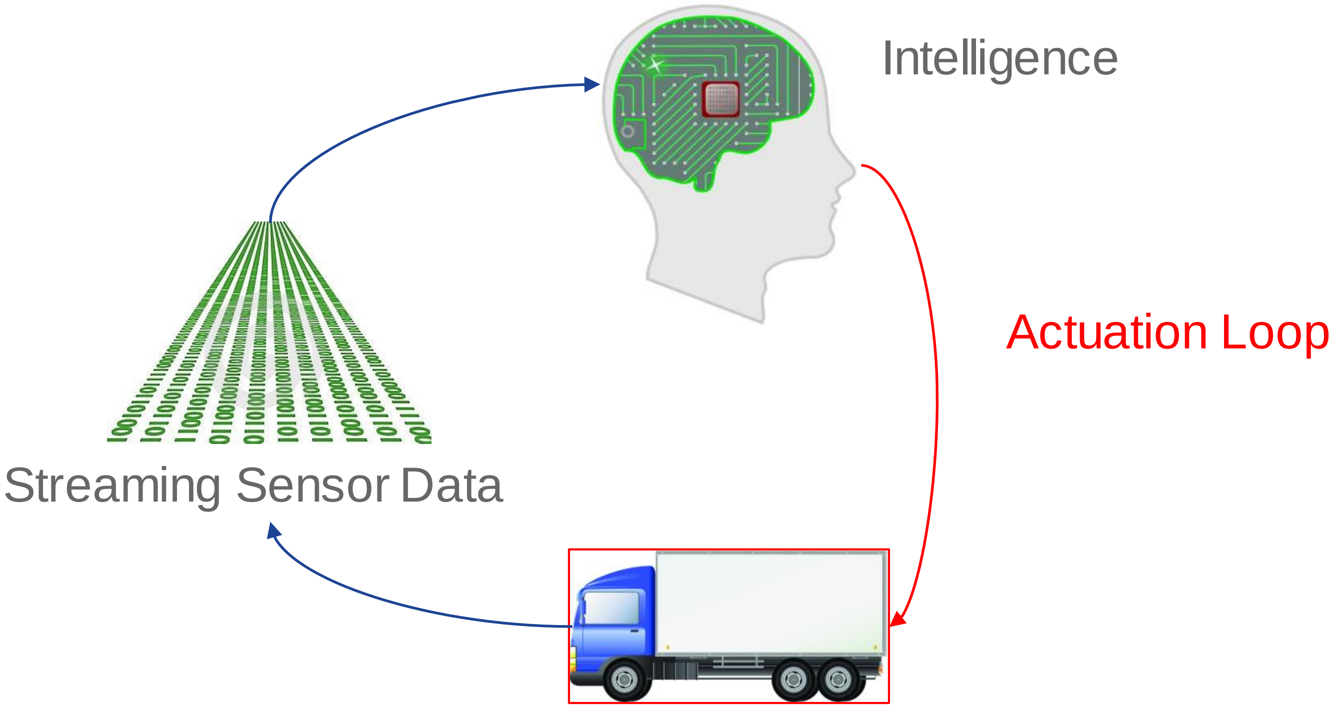
- Labeling Criteria:
  - When tailgating, one brakes too frequently
  - The speeds are usually in 50s and higher
  - Faster reactions/maneuvers
- An expert can detect this
- All of this can be captured locally from Speed and Brake sensors



Tailgating Detection cont'd.

- Collecting speed and brake sensory data from a car
- An expert supervises the labeling
- A Neural Network is trained to classify driving patterns
- The model built by the NN is used to detect tailgating in unlabeled and unseen future data

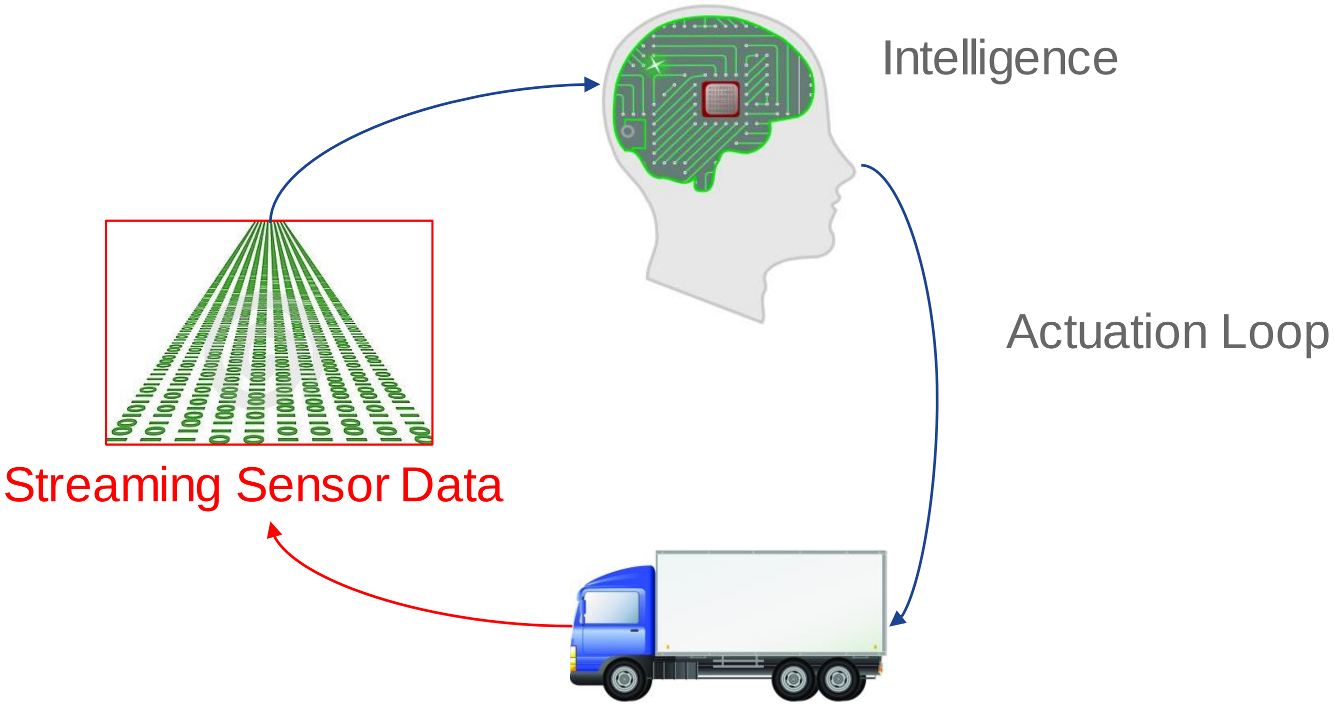
Automation Use Case: Connected Car



Closing the Loop

- Decision making and actuation
- Dangerous driving or Tailgating:
  - Give feedback to driver
  - Inform the insurance company, and/or law enforcement if repeated and necessary
- High fuel consumption:
  - Schedule check up and/or maintenance
- Weather changes that potentially affect tire pressure
  - Inform the driver in advance (for planning)

Automation Use Case: Connected Car



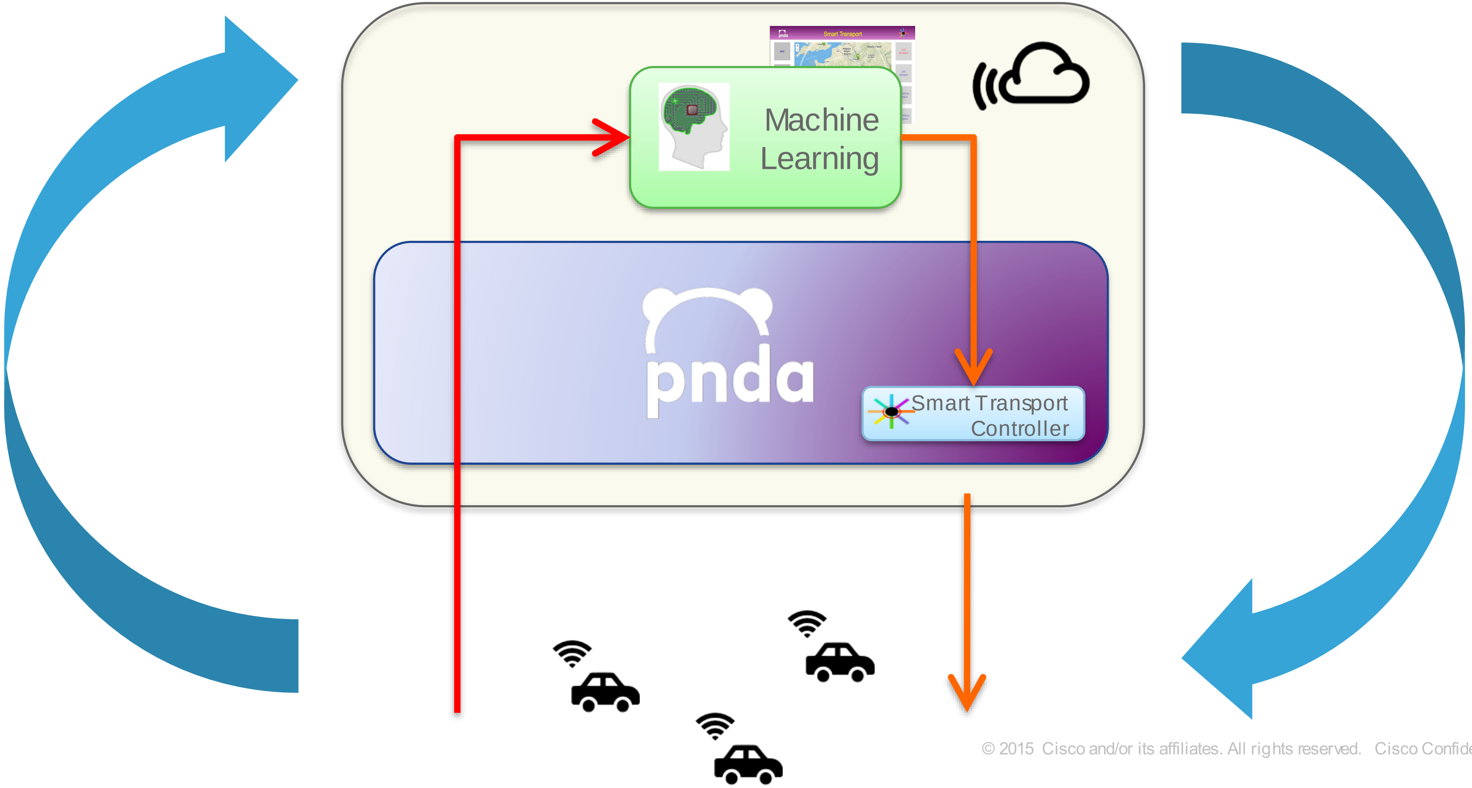
Data Collection

- This is challenging for cars
  - Large amount of data generated
  - Usually slow wireless networks
  - Inherently distributed
  - Scalability is an issue
  - Fast processing is important
  - Security and policy compliance needs to be checked frequently

Scalability

- Built on Scalable and Programmable BIG Data and Policy Control platforms
- Key Scaling Components:
  - Ingestion of vehicle (and contextual) Data into BIG Data platform
  - High performance "Sensor Bus" enabling multiple applications (database, ML, etc.)
  - Real time Machine Learning (ML) detects anomalies and automates distribution of policy data down into the network
  - "Big Data" queries to database
  - Operates in private, hybrid or public clouds → Rapid prototyping, rapid deployment
- Can be adapted for any Sensor/policy-driven Vehicle Network

Machine Learning is an Application



Possible Innovation in Connected Cars

- Correlating driving patterns with fuel consumption/cost, and providing feedback to driver for improvement
- Shaping traffic pattern using connected cars + connected city (traffic lights, etc.)
- Detecting potholes/road deficiencies from on-board car sensors (accelerometer, etc.)
- Detecting impaired driver (under influence, sleepy, etc.) from car sensors (steering, speed, brake, etc.)
- Detecting aggressive drivers and informing other drivers/insurance companies/authorities
- Distributed Machine Learning deployed (for anomaly detection) in the car (as opposed to central cloud) for faster pattern recognition, less data transfer, privacy, etc.

Innovation in Large-scale IoT Deployments

- Pattern recognition
  - Anomaly detection (for security, policy change, etc.)
  - Repetitive patterns and seasonality analysis
  - Trend analysis
- Root cause analysis
- Policy enforcement
- Decision making (closing feedback loop)

Other Potential Use Cases and Data Sources

- Containers (monitoring, anomaly detection, decision making)
  - *data:* CPU, memory, storage usage data
- Routers (pattern recognition, decision making)
  - *data:* TX/RX data rates, NetFlow data
- Home appliances (monitoring, anomaly detection, root cause analysis, policy enforcement)
  - *data:* Energy consumption, network access
- Wearable devices (monitoring, decision making)
  - *data:* Activity data

ML Libraries

- Theano
- TensorFlow
- Scikit Learn
- Caffe
- MLlib

theano

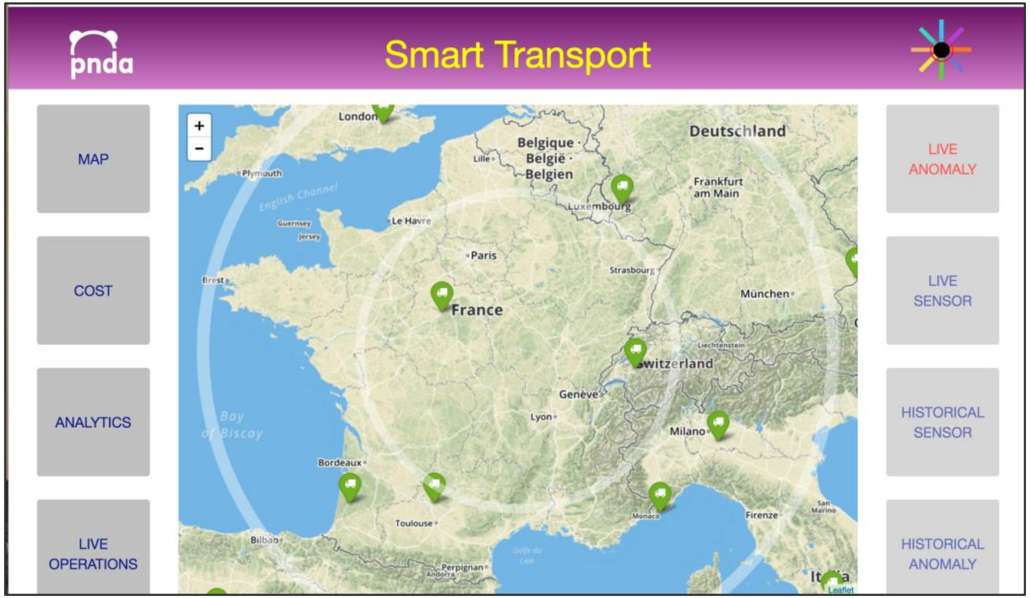
Spark

MLlib

scikit learn

Caffe

Demonstration



https://pndablog.wordpress.com/2017/03/23/smart-transport-analytics-and-machine-learning-using-pnda/

Summary

Summary

- Entering the age of IOT generating BIG DATA
- Connected Cars are the most visible example
- Connected Car Clouds will ingest massive amounts of car data
- Machine Learning easily detects complicated data patterns regardless of scale
- Machine Learning remove humans from the loop
- Machine Learning makes our life easier

Any follow ups/ questions send to:

- Dr. Mehdi Nikkhah, mnikkha2@cisco.com
- Patricia Sampedro Garcia, psampedr@cisco.com
- Chris Metz, chmetz@cisco.com

References

- Machine Learning, Ebert, C., Louridas, P. IEEE Software (Volume: 33, Issue: 5, Sept.-Oct. 2016)
- https://blogs.cisco.com/tag/machine-learning
- Python Machine Learning, Sebastian Raschka
- https://www.udemy.com/scala-and-spark-for-big-data-and-machine-learning/
- Cisco Blog on Machine Learning, Mehdi Nikkhah

Questions

?