Deep Learning for Urban Transportation

A MIT Seminar 11.s938J/11.S196J Fall 2019

Peyman Noursalehi, Shenhao Wang and Jinhua Zhao
Massachusetts Institute of Technology

What's changing in Transportation?

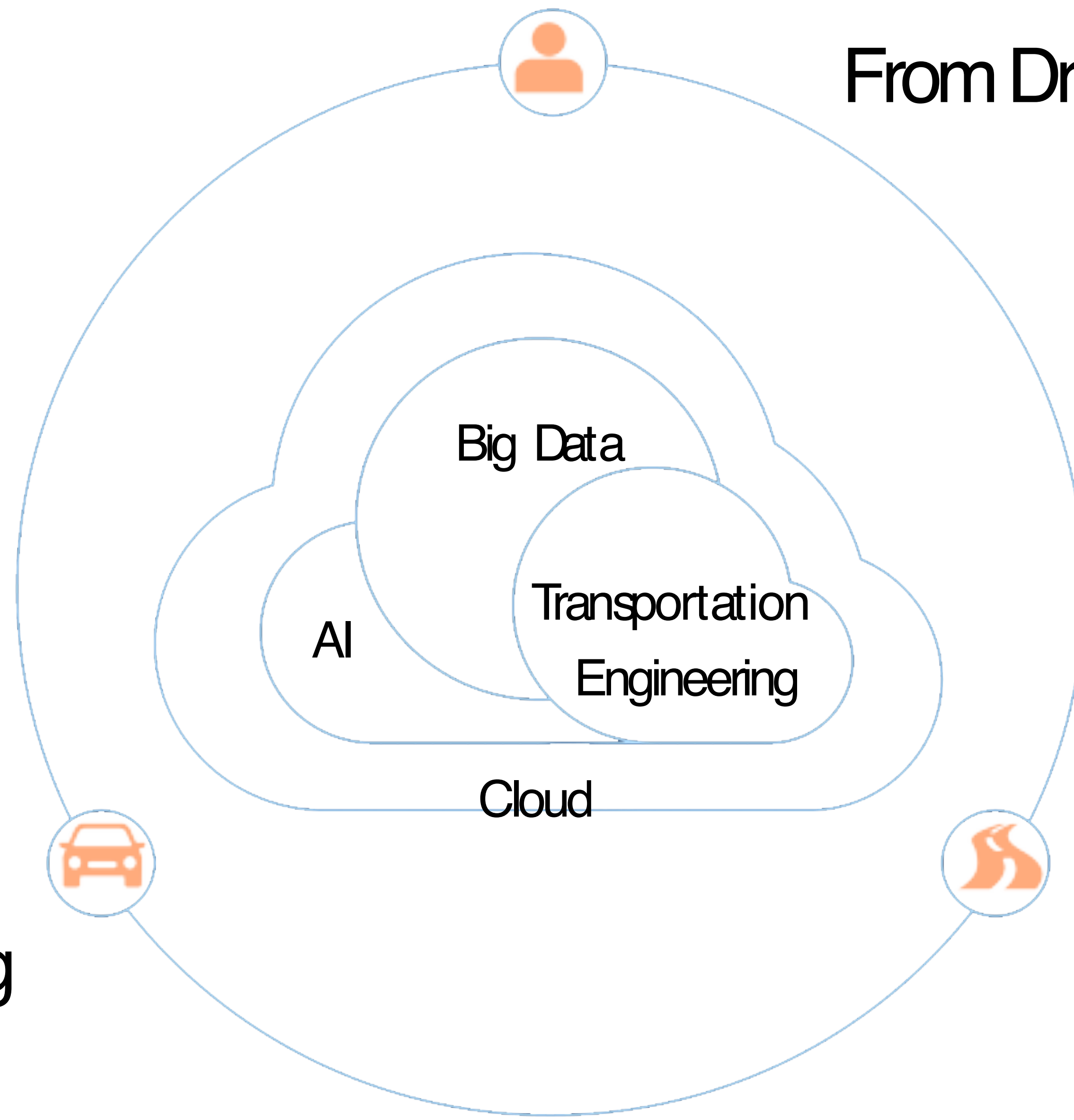
What’s changing in transportation?

Technology	Data	Value
• Automation • Electrification • 5G/Connected • Shared economy • …	• Ubiquitous sensing • AI / computing • Cybersecurity • Clearing house • …	• Climate change • Future of work • Public health • Social justice • Urban livability • …

People, Vehicles and Infrastructure

Smart Travelers

From Drive Alone to Ride Sharing



Smart Vehicles

From Human Driving to Autonomous Driving

Smart Infrastructure

From Independent Systems to Cooperative Vehicle-Highway Systems

Tech Uncertainty —> System Uncertainty

- Ownership model or access Model
- Single use or shared use
- Integrate with public transit or not
- Rational pricing structure or not
- Coordinated with land use or not
- Rational regulatory framework or not
- …

Multimodal Transportation

Data


The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

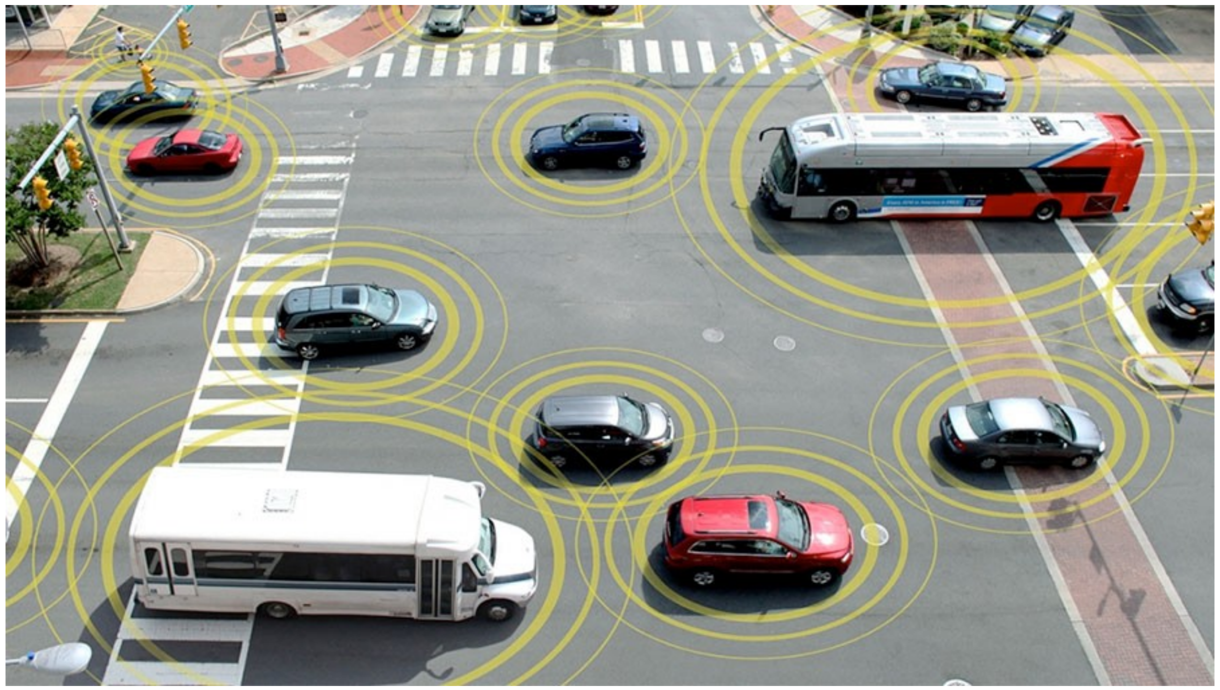
Data Resource

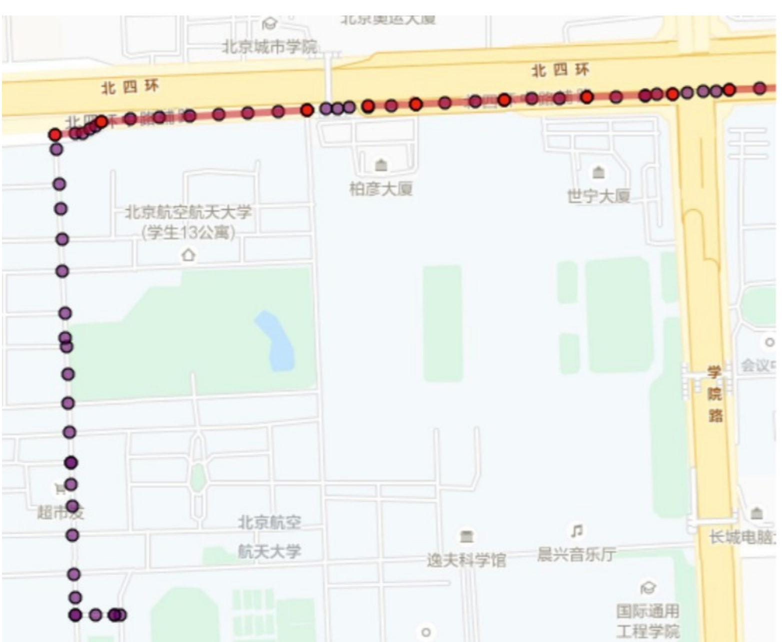
- Location data, Trajectory data
- Transaction data
- Profile data
- Sensors: multimedia data



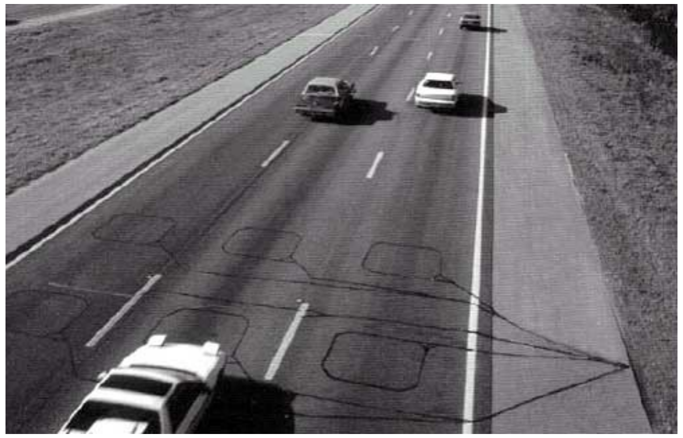
- Cross-platform identification

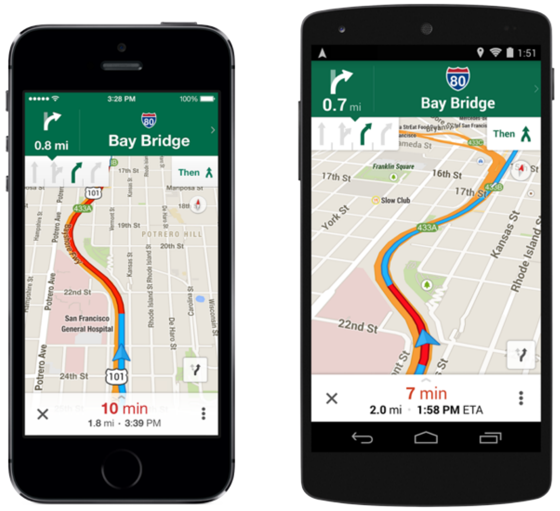

Location Data and Floating-Car Trajectory





Sensors: Loop detector, camera, microphone, mobile sensors . . .





Open Datasets

KDD2017

KDD Cup 2017

Highway Tollgates Traffic Flow Prediction

UBER

Uber Movement

NYC OpenData

Public Data

U.S. Department of Transportation
Federal Highway Administration
Federal Highway Administration Research and Technology
Coordinating, Developing, and Delivering Highway Transportation Innovations

Federal Highway Administration

Next Generation Simulation (NGSIM) Program

DiDi

GAIA Open Dataset

Trajectory Data

- LTDS / CRM
- Oyster / Octopus
- iBus
- Mobile CDR / Wifi / RFID / Bluetooth
- Tracking Apps: MOVES / FMS / Strava
- Mobile App
- Text: Incidence Log, Customer Feedback, Twitter, Facebook, What’s App…

Applying Machine Learning Techniques to Transportation Surveys

Jane Shepherd, Westat
Marcelo Simas, Westat
Anthony Fucci, Westat
Alexander Cates, Westat

Background

- Household Travel Surveys
  - Collect socio-economic and demographic data about households and individual members
  - Collect a travel diary for 1-2 days
    - Describe the how, why, when, and where of each place visited on the assigned travel day(s)
  - Recently deployed smartphone-based surveys
    - Geolocation – Auto-detects trip start/stops using geofences
    - Travel capture – GPS data informs arrival and departure times
    - Prompted recall
  - Past surveys
    - Asheville, Fairbanks, Albuquerque, South Jersey, Las Vegas, Michigan, Billings, NHTS
  - Present / future surveys
    - Chicago, Maryland, Laredo

How do we use machine learning?

1. Coding open-end responses using Natural Language Processing and Random Forest models.

2. Ascertaining Industry and Occupation in real time using Natural Language Processing and Vector Space models.

3. Determining place validity and predicting travel attributes using GPS and Accelerometer-derived features to train Random Forest models.

Coding Open-End Responses

- Problem
  - NHTS yielded over 180,000 open-ended responses
  - Around 52,000 of these belonged to the “Trip Purpose” question
- Traditional Solution
  - Analyst attempts to up-code each response by hand
  - Average 15 sec / response = ~ 750 hours
- ML Solution
  - Analyst up-codes a sample of responses.
  - Treat the sample as labeled training data to be modeled

Breadth of Coverage (User, Time, Space): Oyster + Mobile CDR

?

Statistical Data Fusion via Machine Learning

Depth of Attributes (Semantics)
LTDS + CRM + Tracking

?

Statistical Data Fusion via Machine Learning

Multi-modal Data Fusion and Analytics

Multimodal Travel Information Platform

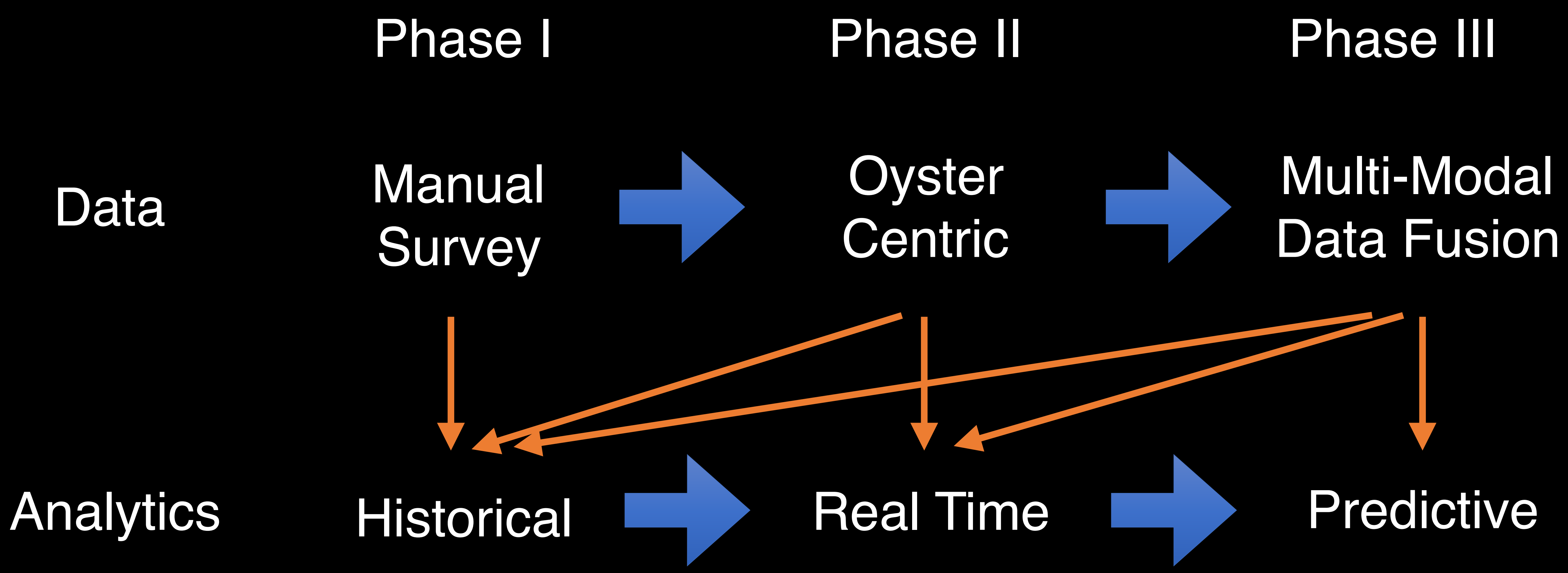
- Oyster
- Mobile CDR / Wifi / RFID / Bluetooth
- LTDS / CRM
- Tracking Apps: MOVES / FMS

+

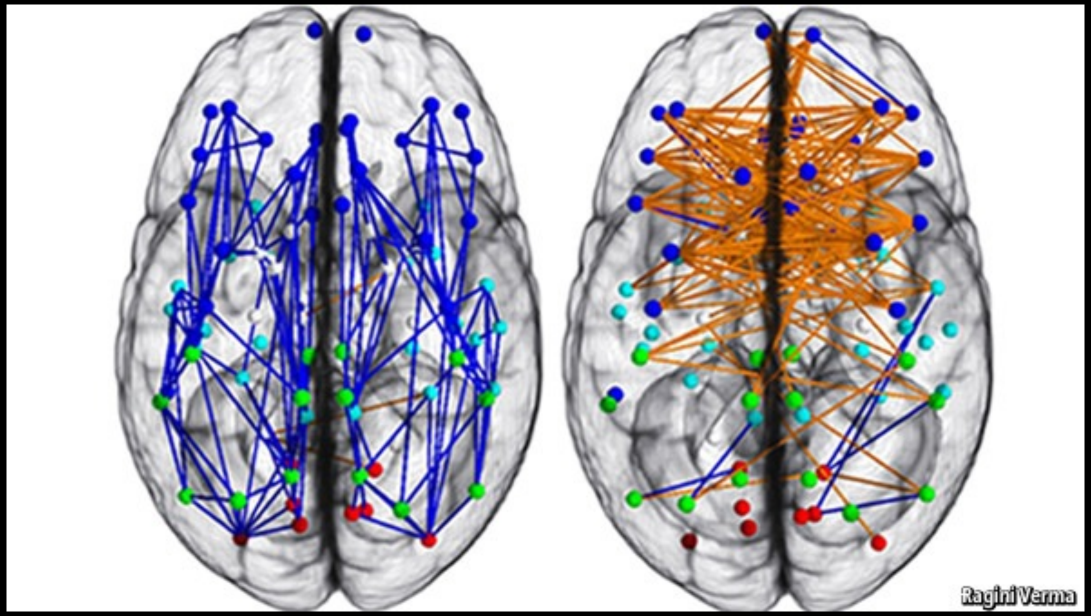
Statistical Data Fusion

- Discrete Choice Modeling
- Clustering: Unsupervised Learning
- Machine Learning: Neural Network / Random Forest /Support Vector Machine …

TfL Data Analytics




Car
or
Positive Social
Status
Bus
or
Negative Social Status
Implicit Association
Test


Ragini Verma

SP + RP

Multitask Learning Deep Neural Networks to Combine Revealed and Stated Preference Data

Shenhao Wang
Qingyi Wang
Jinhua Zhao

Massachusetts Institute of Technology

ICMC 2019

Background

RP+SP as a Classical Question

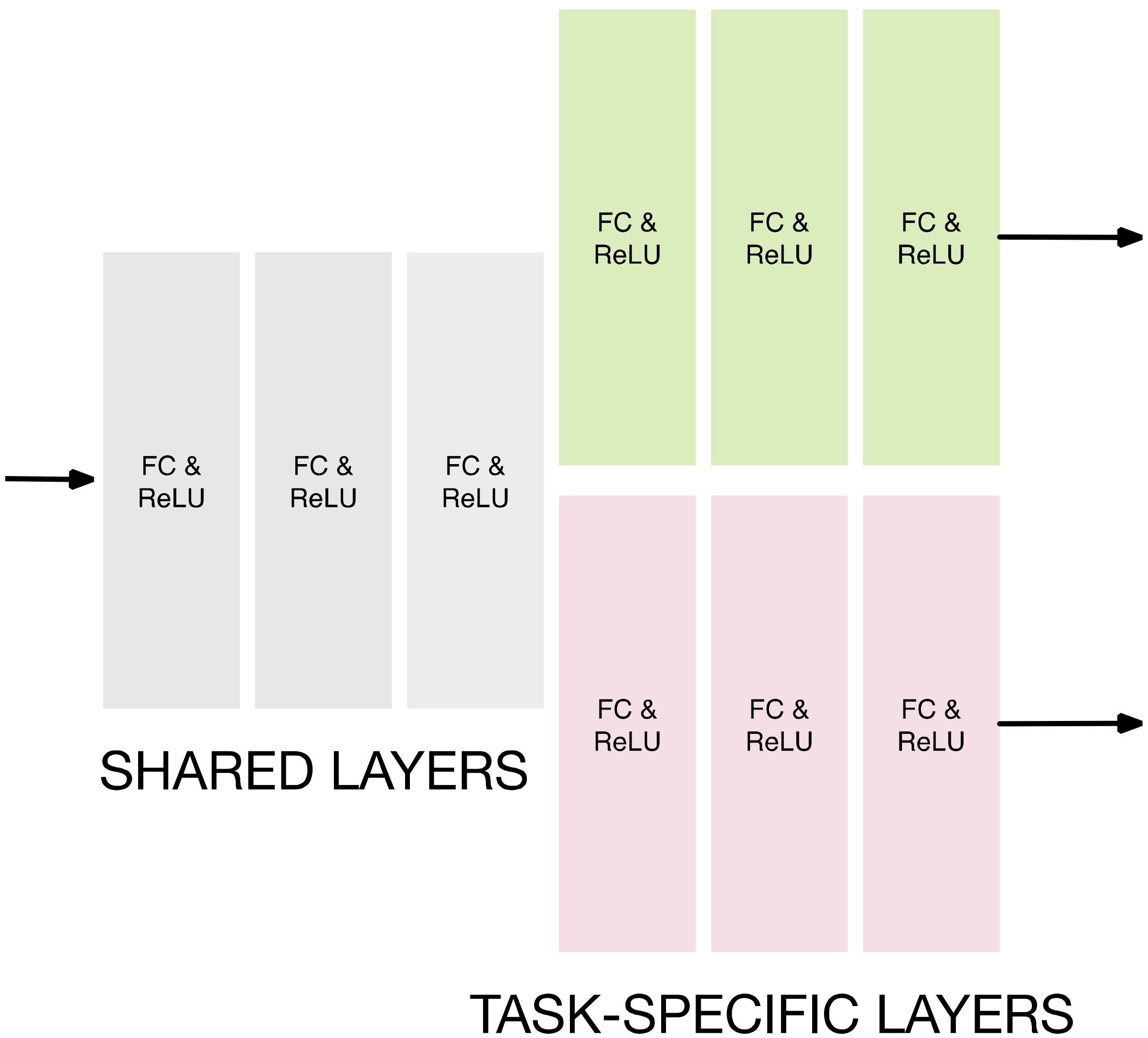
- Pros and cons of RP and SP (Ben-Akiva et al., 1994; Hausman et al., 1998)
- Joint RP+SP (Ben-Akiva et al., 1994; Hensher and Bradley, 1993; Polydoropoulou and Ben-Akiva, 1994)
- Nested logit model as one classical method (Hensher and Bradley, 1993; Louviere et al. 1999)

MTLDNNs as a New Method

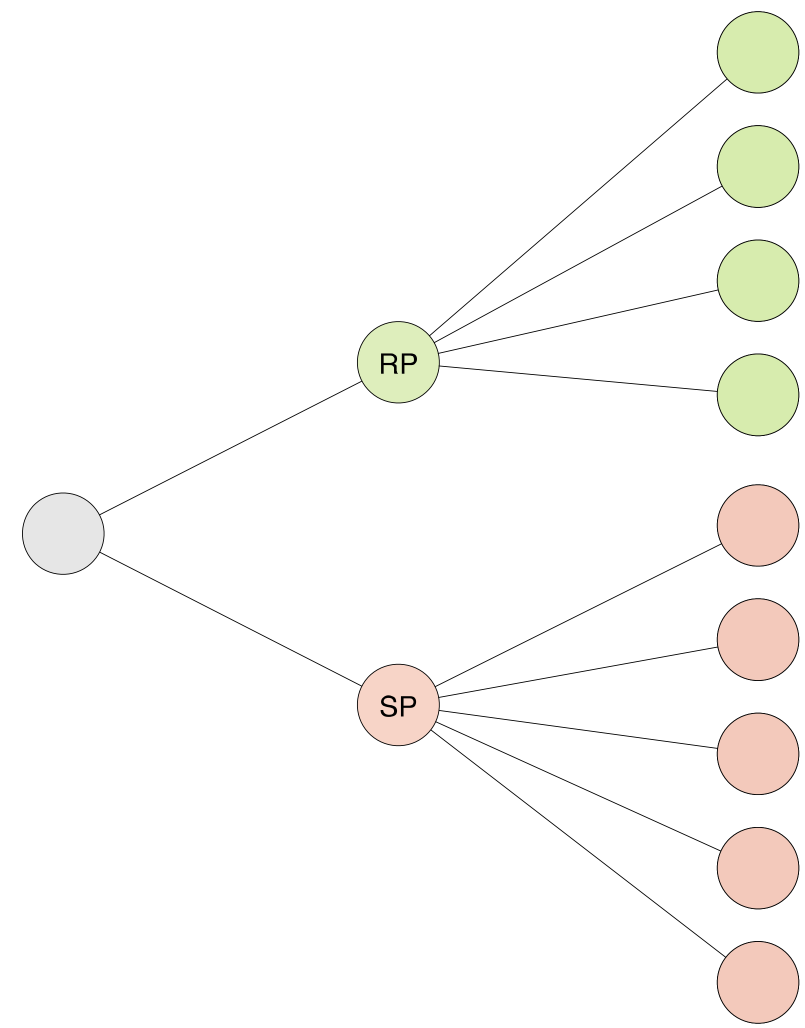
- A multitask learning perspective
- ”Multiple tasks arise naturally…” (Caruana, 1997)
- Wide applications: NLP (Collobert and Weston, 2008; Hashimoto et al. 2016); healthcare drug discovery (Ramsundar et al. 2015); etc.

An Intriguing Question: MTLDNNs and NLs

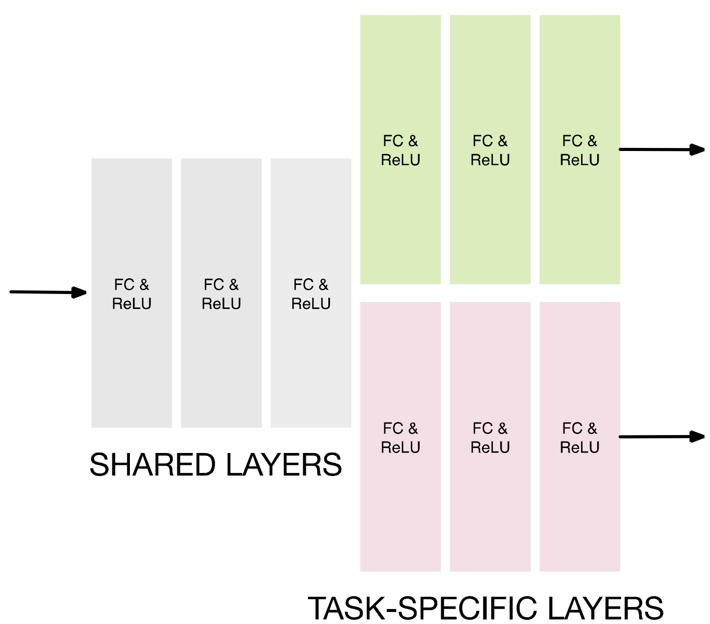
MTLDNN Visualization



NL Visualization



Formulation of MTLDNNs



Empirical Risk Minimization

$$\min_{w_r, w_s, w_0, T} R(X, Y; w_r, w_s, w_0, T; c_H) =$$

1 Cross-Entropy Loss

$$\min_{w_r, w_s, w_0, T} \Bigg\{ \boxed{ -\frac{1}{N_r} \sum_{i=1}^{N_r} \sum_{k_r=1}^{K_r} y_{k_r} \log P(y_{k_r, i}; w_r, w_0; c_H) - \frac{\lambda_0}{N_s} \sum_{t=1}^{N_s} \sum_{k_s=1}^{K_s} y_{k_s} \log P(y_{k_s, t}; w_r, w_0, T; c_H) } + \boxed{ \lambda_1 ||w_0||_2^2 + \lambda_2 ||w_s||_2^2 + \lambda_3 ||w_s - w_r||_2^2 } \Bigg\}$$

2 Regularizations: Scale Controls

MTLDNNs are More Generic than NLs.

MTLDNNs

1. Automatic feature learning
2. "Soft" constraints to describe the similarities between RP and SP
   – Architectural design (e.g. # of shared vs. task-specific layers)
   – Regularizations (e.g. λ_3)
3. Possible to specify the model with a massive number of tasks

NLs

1. Handcrafted feature learning
2. "Hard" constraints to describe the similarities between RP and SP
   – Shared vs. task-specific parameters (e.g. β_r vs. β_s)
3. Hard to specify the model with a massive number of tasks

Deep Learning for Urban Mobility

Part I: Introduction and DNN Basics
- Sep 10 Introduction: Deep Learning Meets Transportation

Part II: Passenger and Traffic Flow Prediction
- Sep 17 Demand prediction and Deep learning basics
- Sep 24 Advanced modeling techniques: ConvLSTM, Attention, individualized predictions
- Oct 1 Guest Lecture Prof. Justin Dauwels
- Oct 8 Advanced Applications: Generative models (VAE), graph embeddings,

Part III: DNN and Demand Analysis
- Oct 22 DNN and Discrete Choice 1
- Oct 29 DNN and Discrete Choice 2
- Nov 5 DNN and Prospect Theory

Part IV: Reinforcement Learning and Control in Transportation
- Nov 12 RL and Control Part 1
- Nov 19 RL and Control Part 2
- Nov 26 Guest Lecture Prof. Cathy Wu

Part V: FAT and Summary
- Dec 3 Fairness, Accountability and Transparency
- Dec 10 Student Presentation

Style of the Class

- Balance: transportation ~ machine learning
- Seminar: a bit ad hoc, a survey
- Credit: 6, p/d/f
- Expectation of students
  - Class participation
  - Term project

Applications

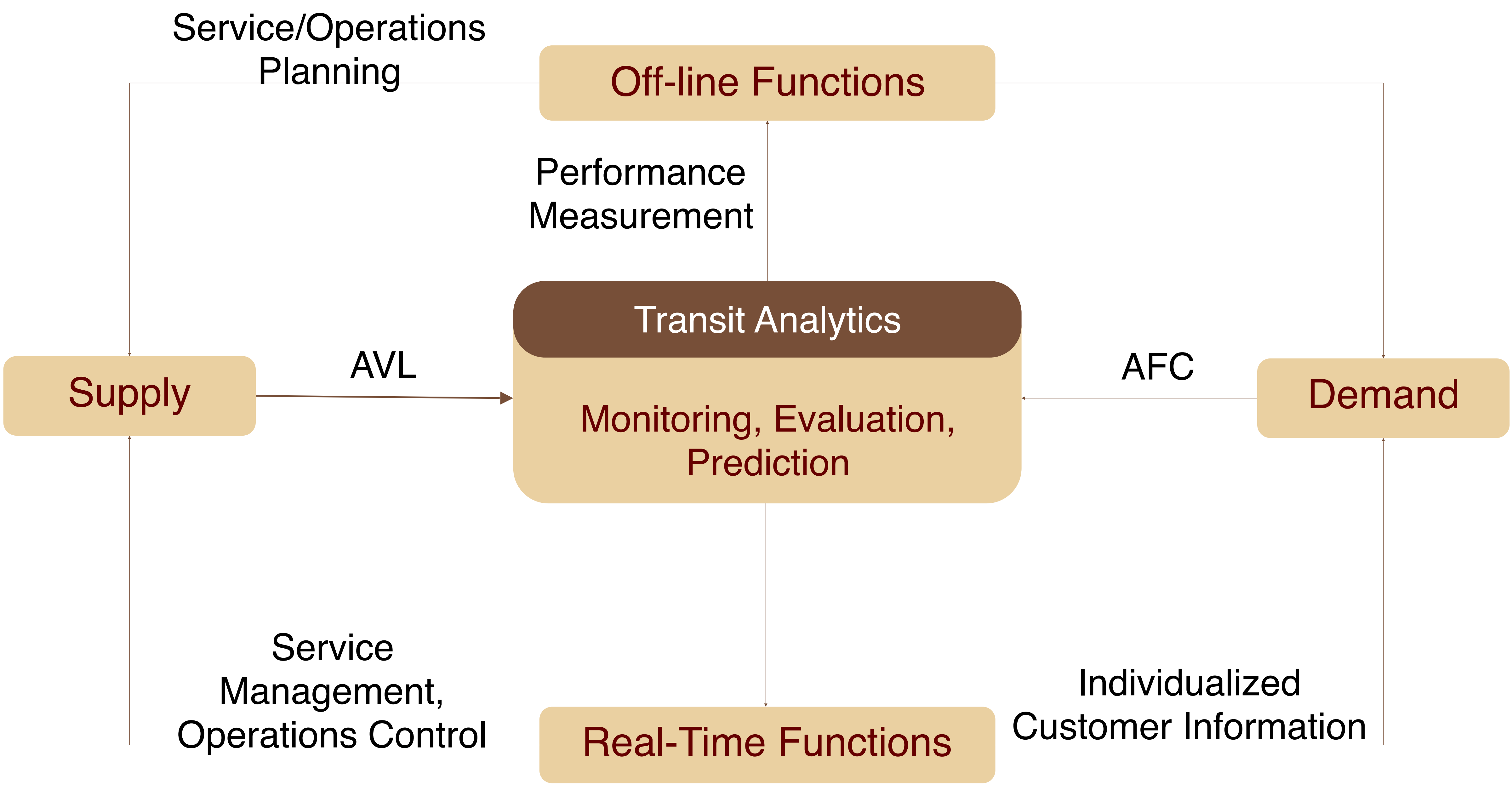
ML for transit


TEZENIS
TEZENIS
TEZENIS
TEZENIS
UNDERGROUND
at platform level
MORE LINES ADDED
SALE

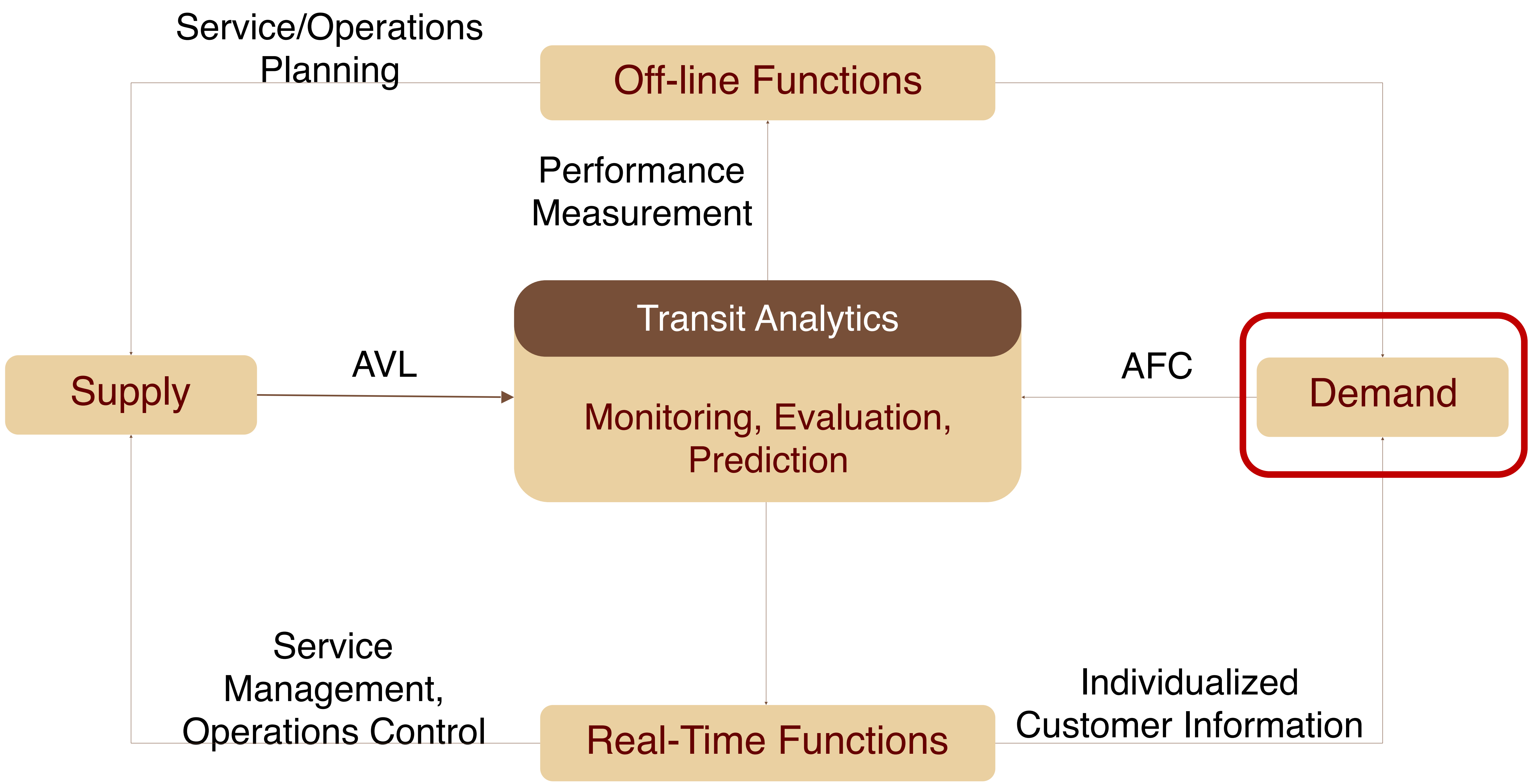
Challenges

- Near capacity operation
- Crowding (train, platfor
- Passengers left behind
- Safety concerns
- Disruptions

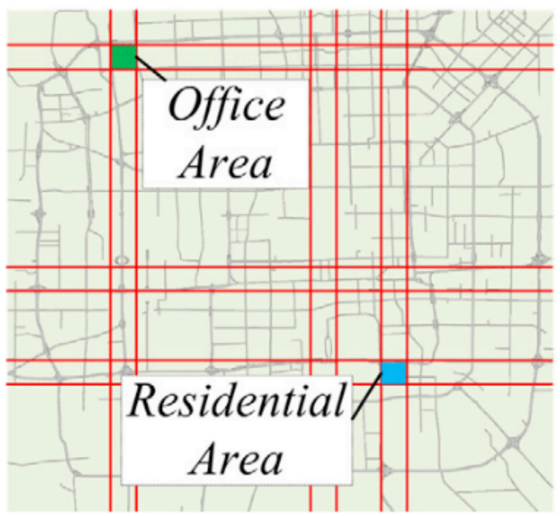
Overall framework



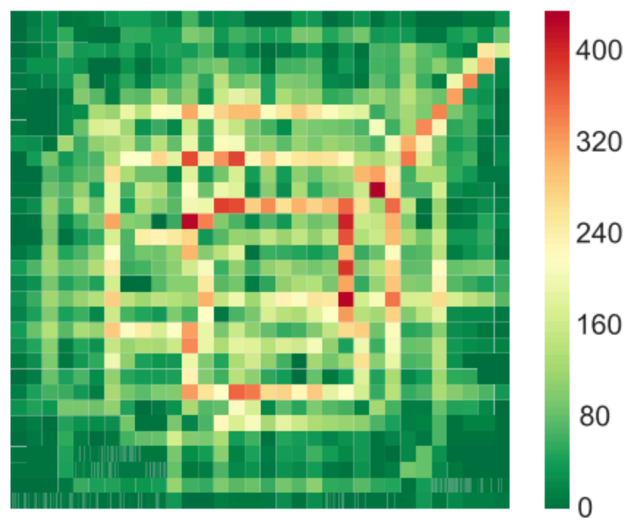
Overall framework



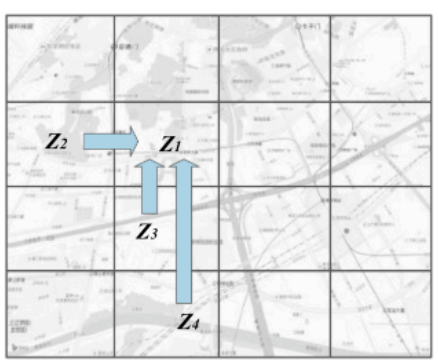
Aggregate demand prediction

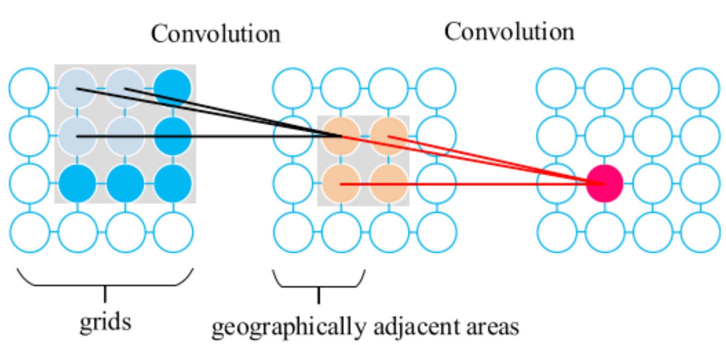


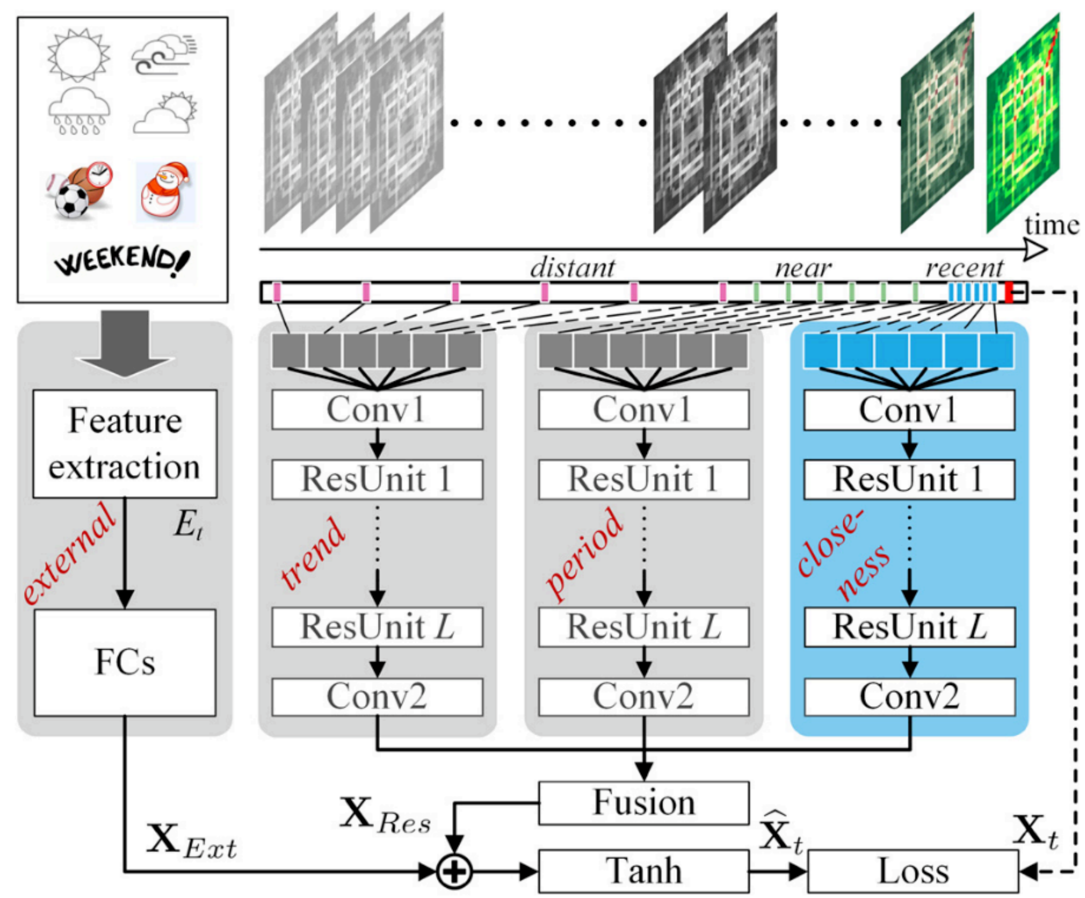
(a) Map segmentation



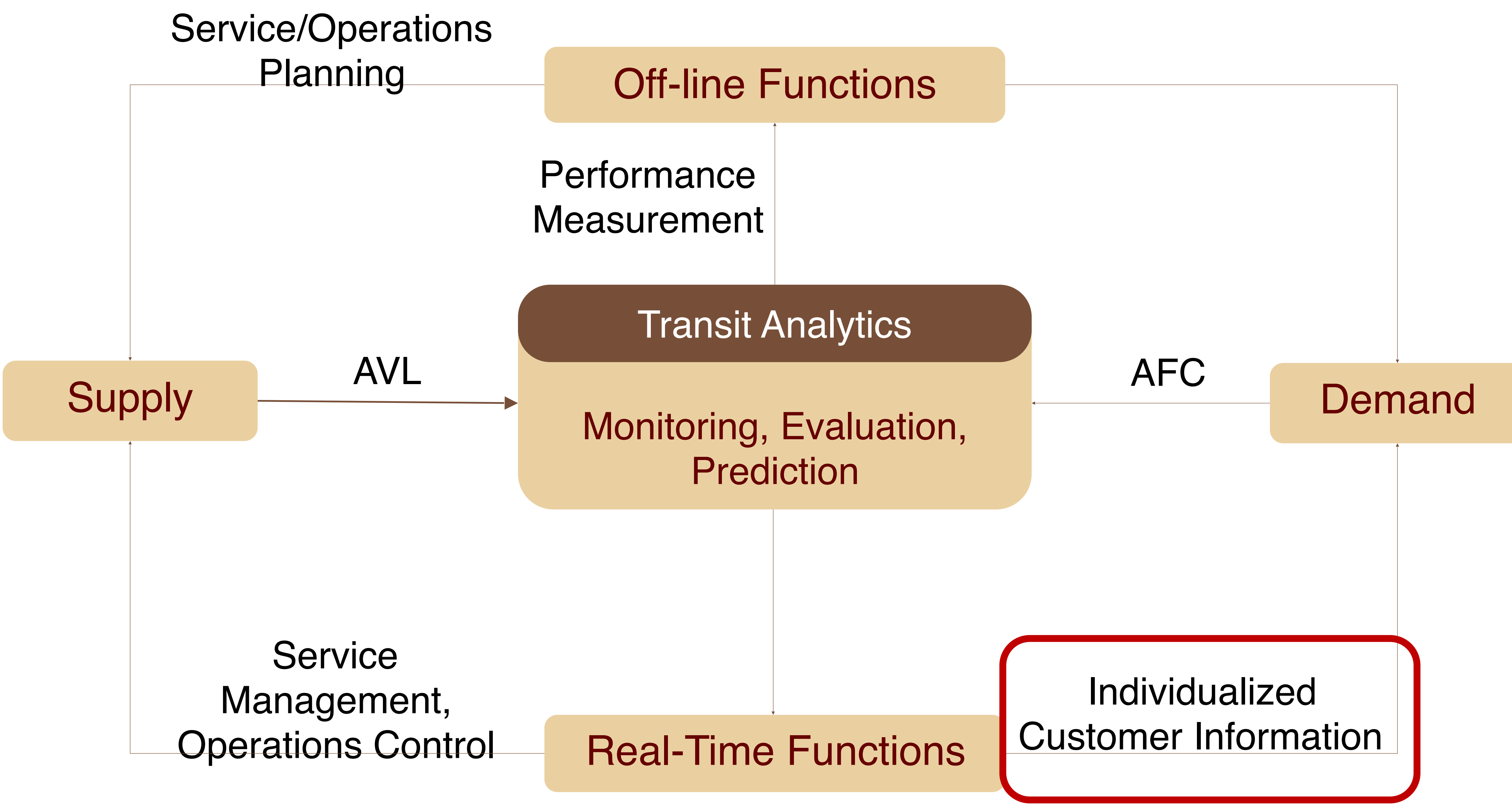
(b) Inflow matrix





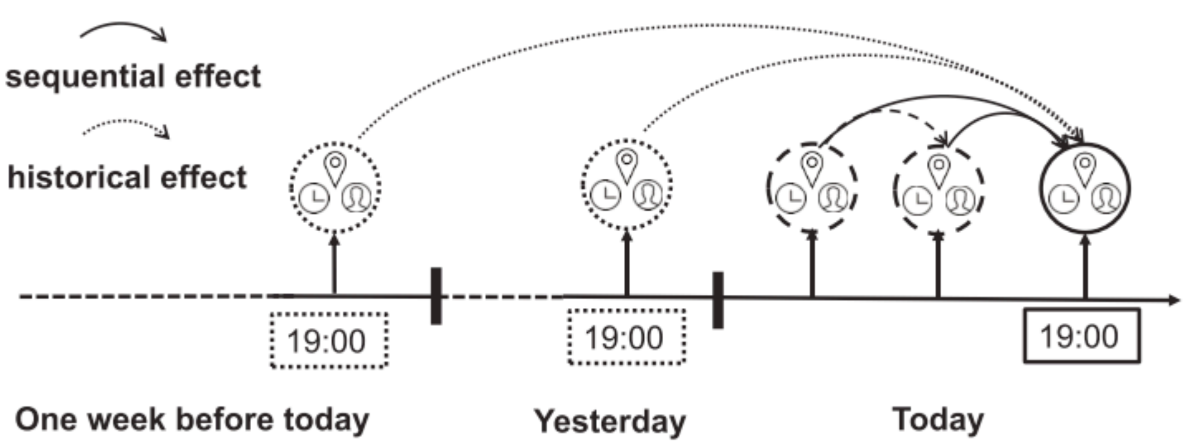


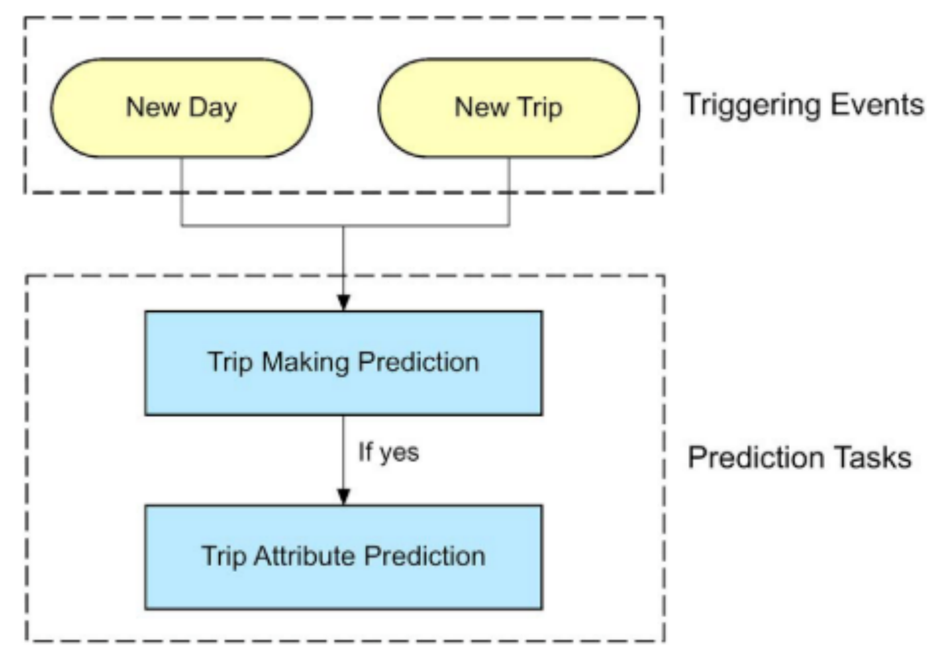
Overall framework



Individual mobility prediction

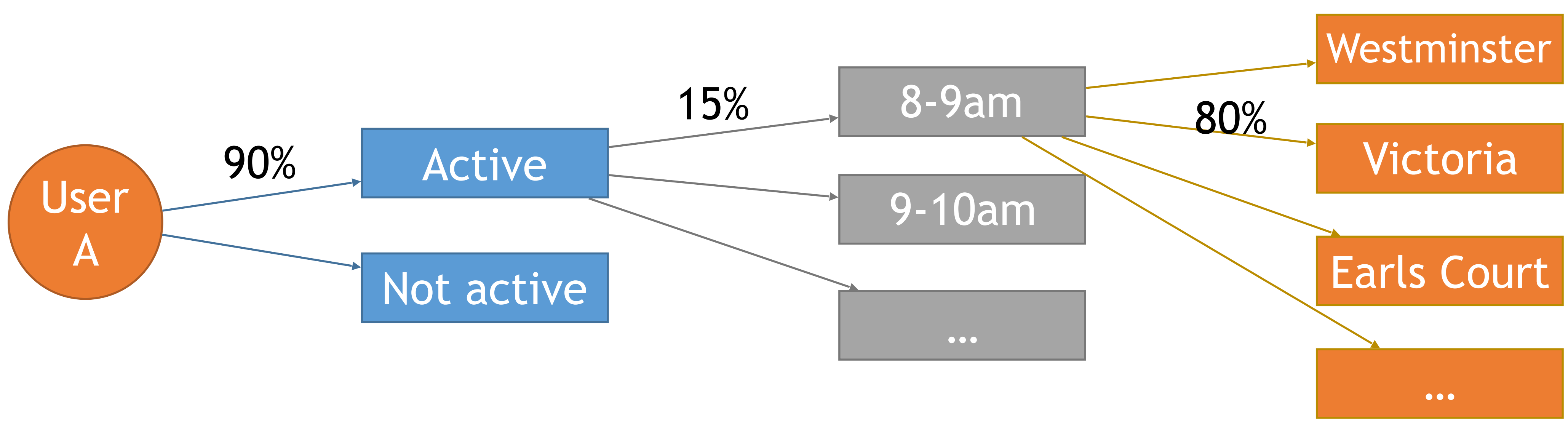
- When an incidence/delay occurs, send travel alerts to passengers who are likely to be affected



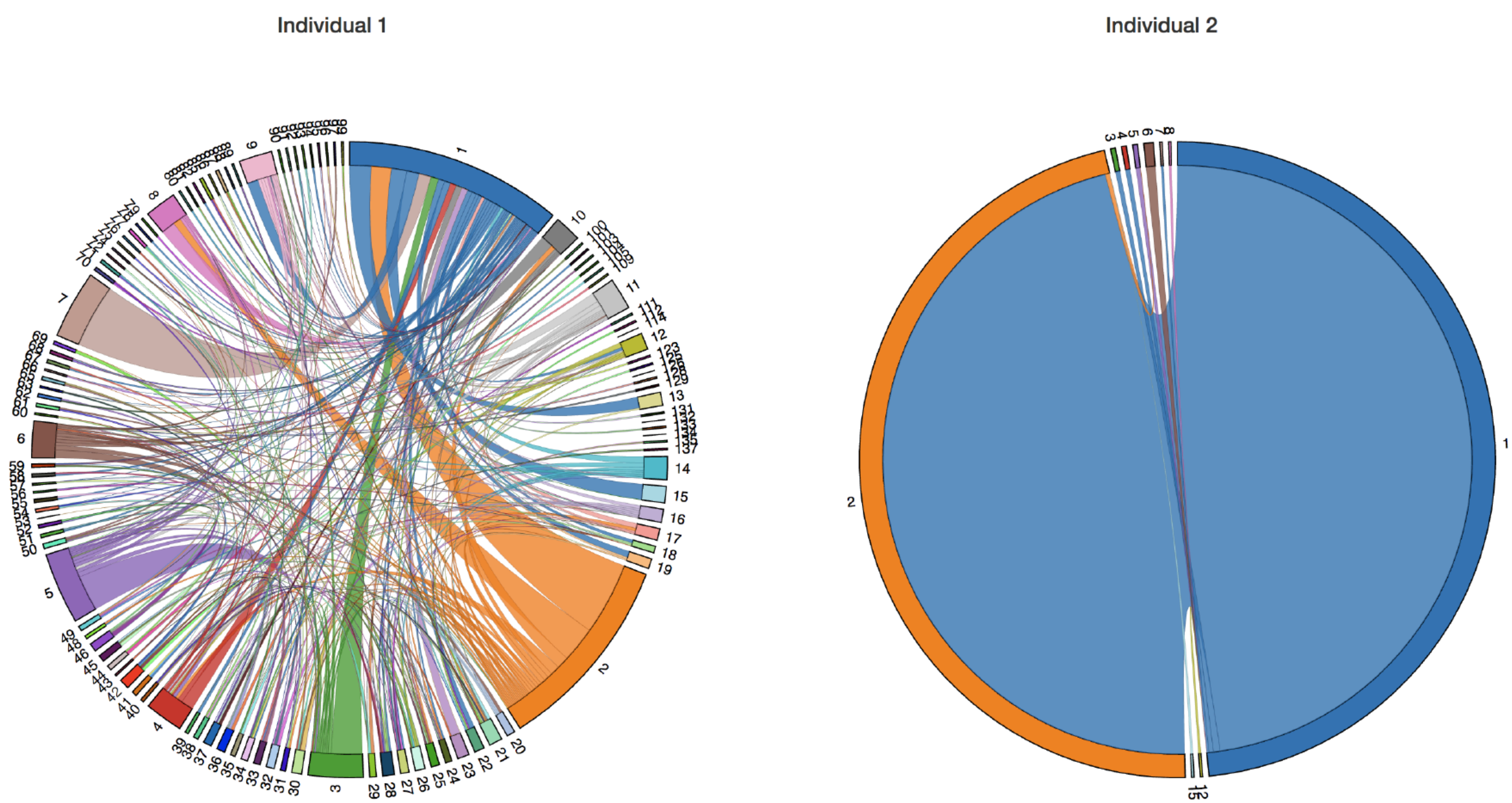


Individual mobility prediction

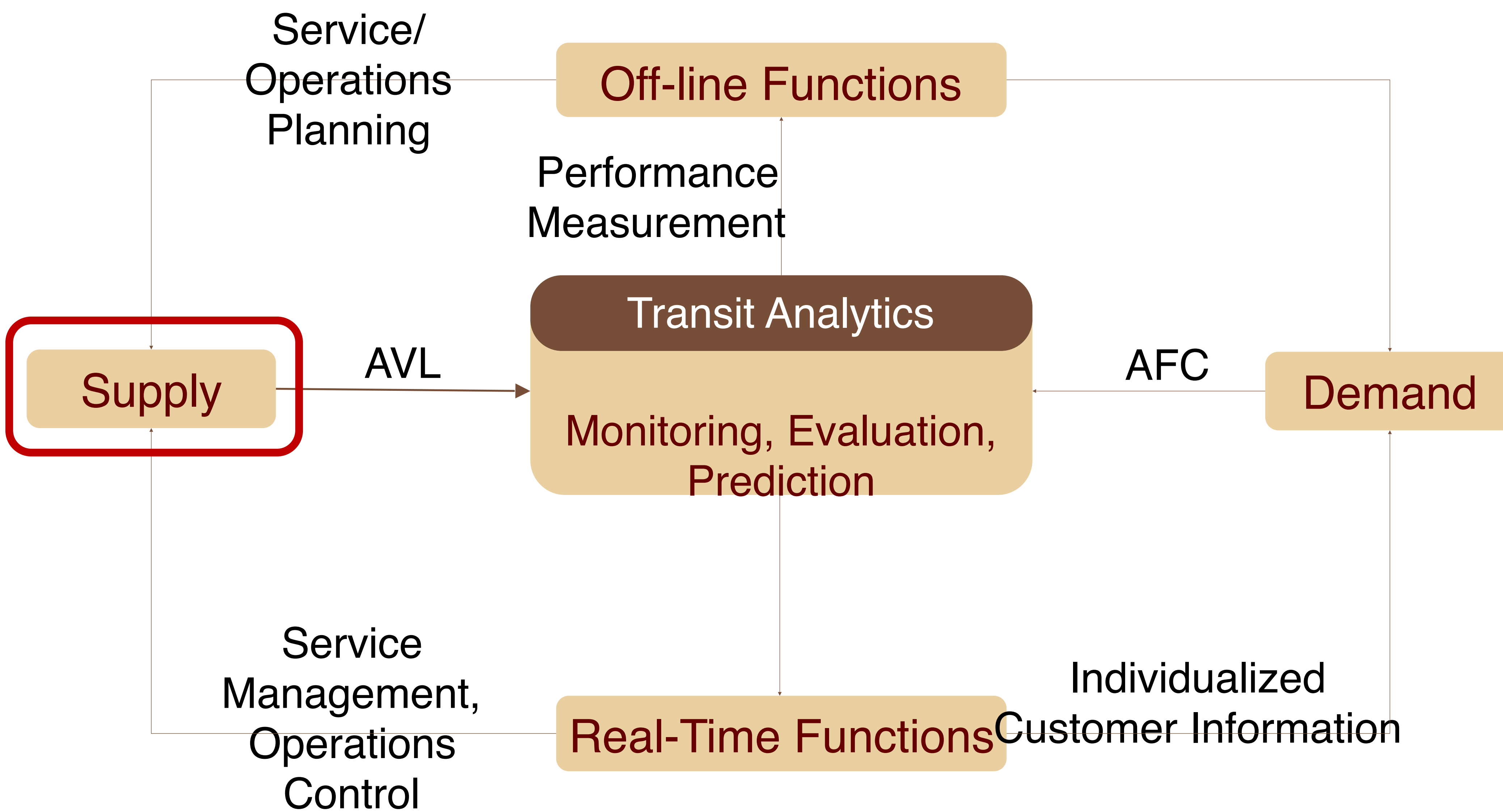
- Example: at 8 am, an incident occurs in Victoria Station, and will likely affect the service for the next hour



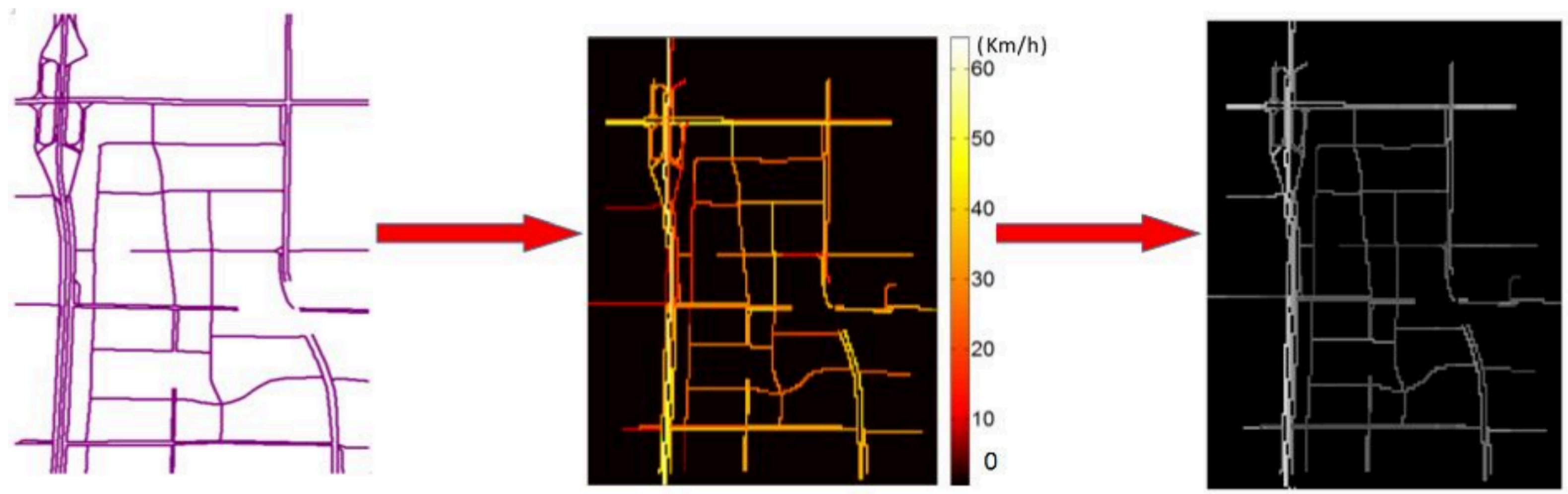
Individual mobility prediction

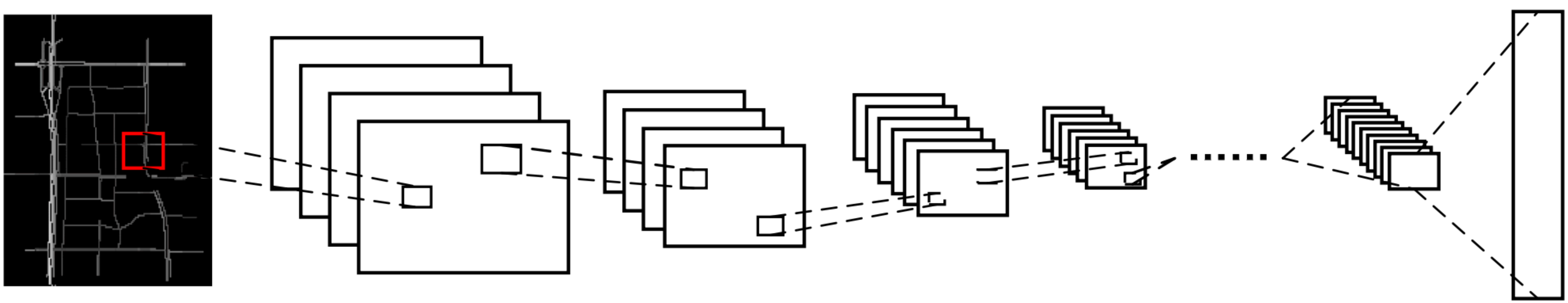


Overall framework

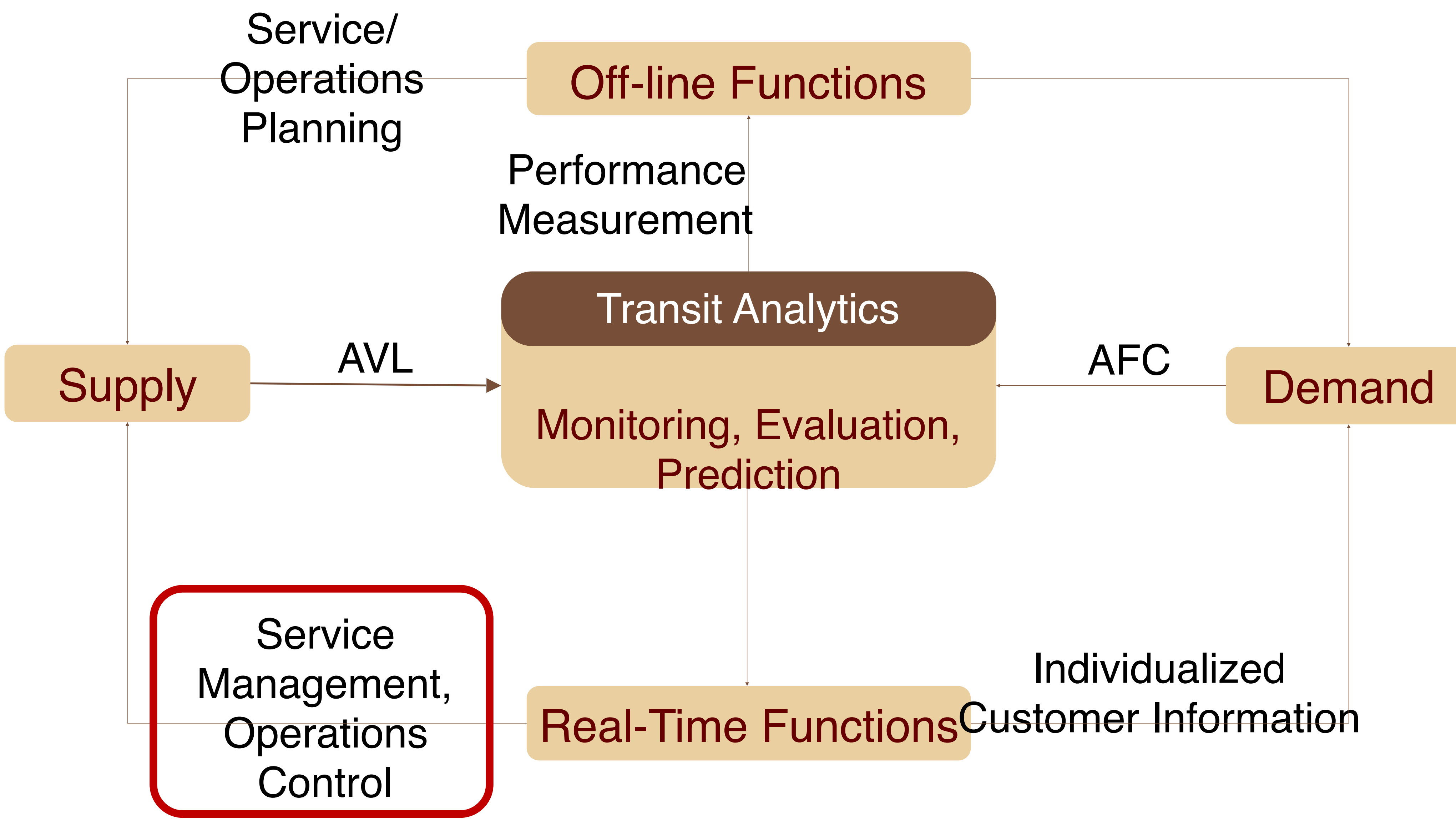


Traffic prediction





Overall framework

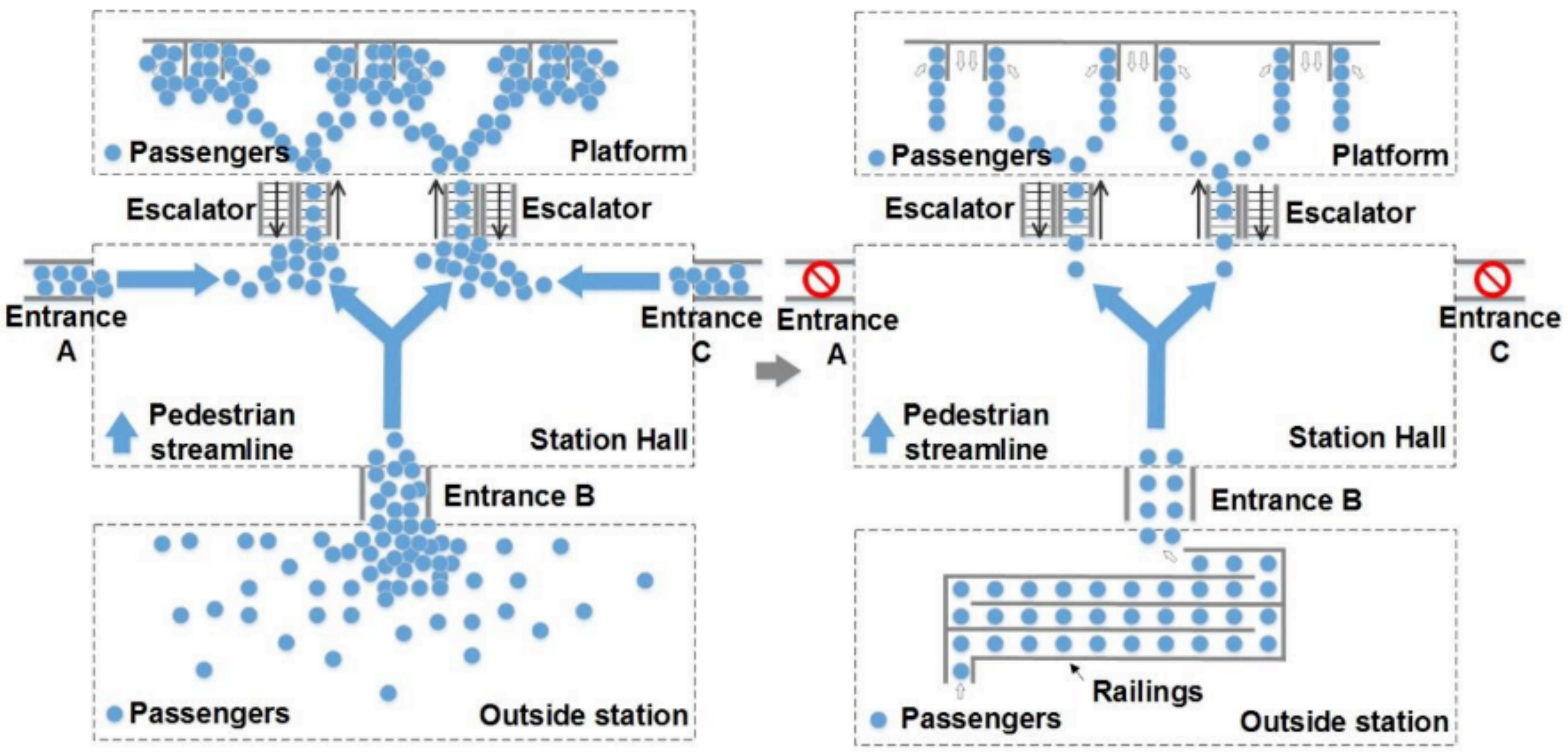


Opportunities

- Proactive crowd management

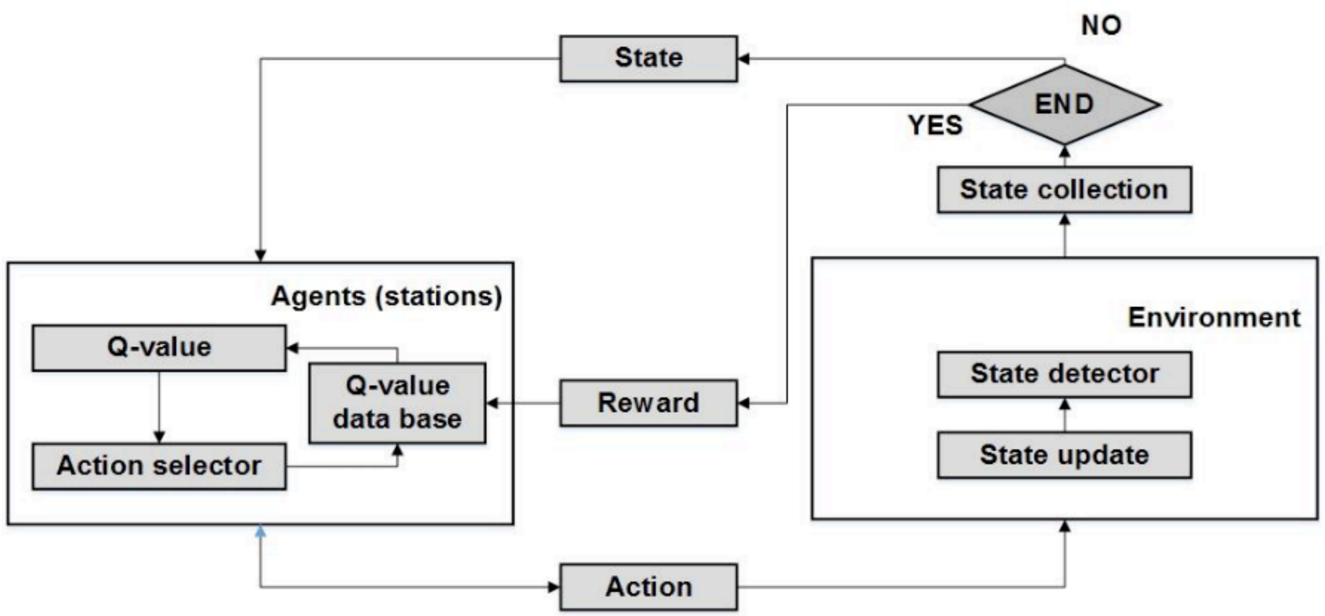
Proactive crowd management

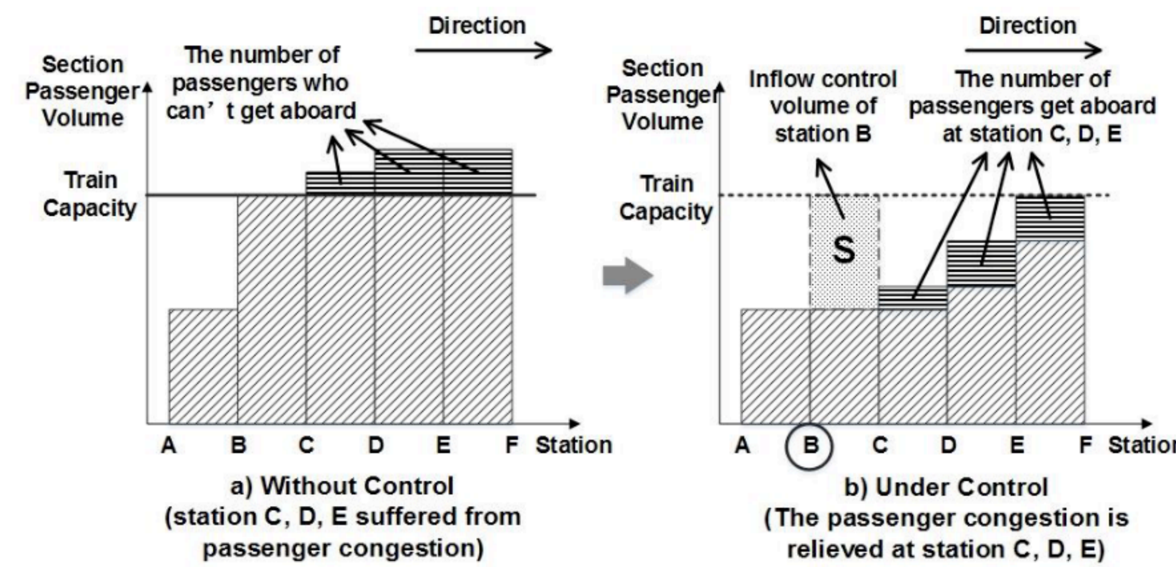
- Example: Beijing Metro



a) Without Control b) Under Control

Proactive crowd management





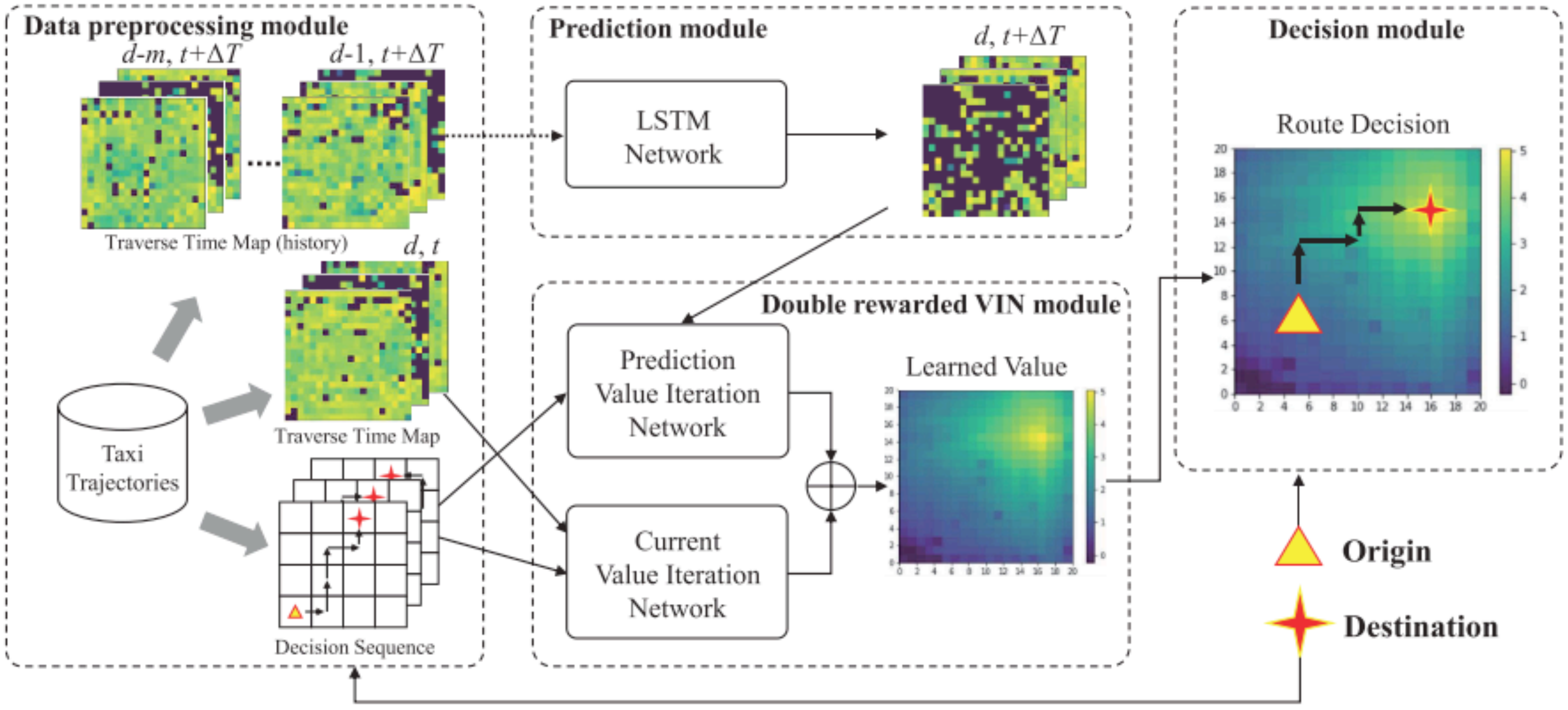
ML for ride hailing services

Ride hailing systems

- Short-term demand forecasting is crucial for operation
- Examples: Uber, Didi, and Lyft
- Applications:
  - Dispatching efficiency
    - Driver to zone assignment
  - Adaptive pricing
  - Rebalancing
  - Passenger information

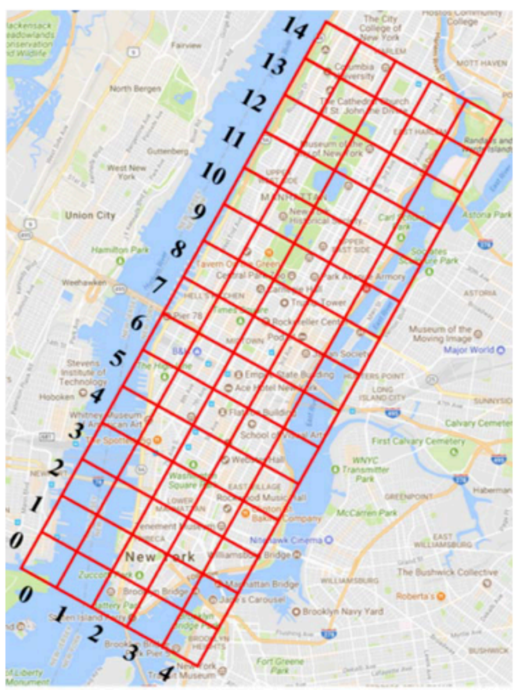
Ride hailing systems with AV

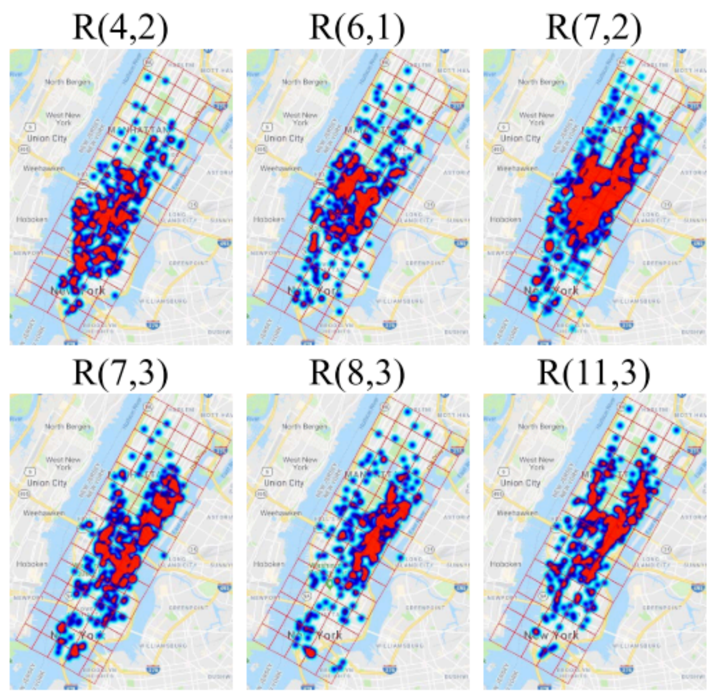
- RL for centrally controlled taxi operation

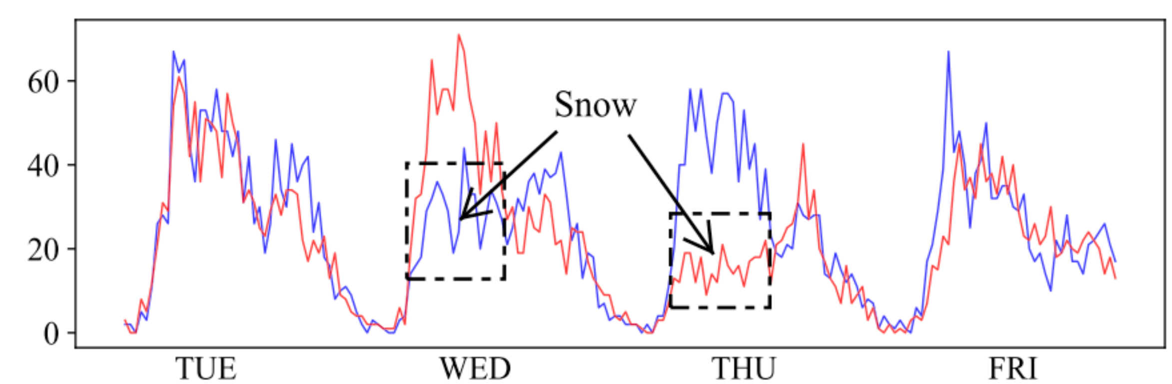


Large-scale Demand prediction

- Origin-destination demand for taxi in NYC

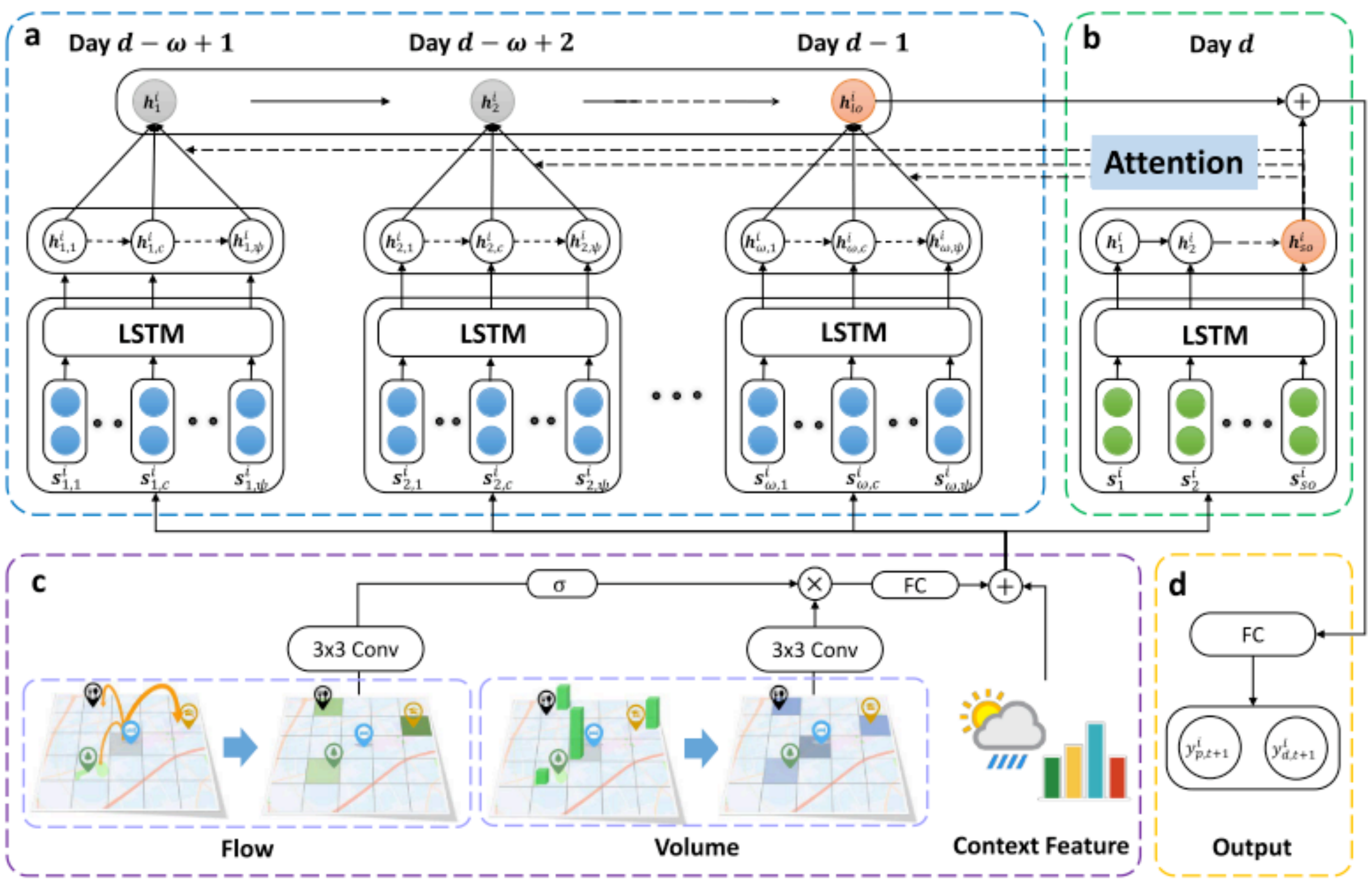






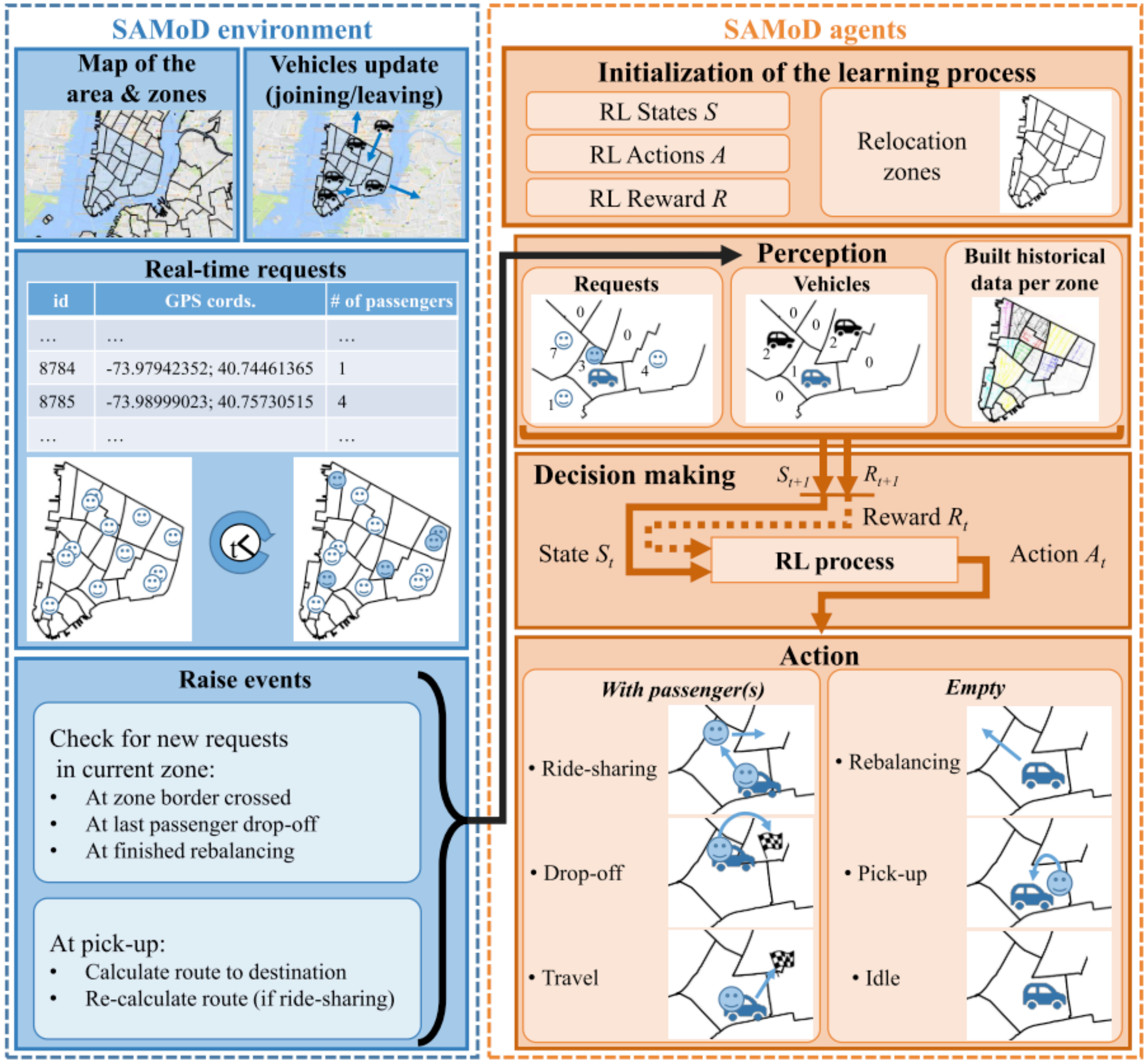
Large-scale Demand prediction

- Origin-destination demand for taxi in NYC



Ride hailing systems with AV

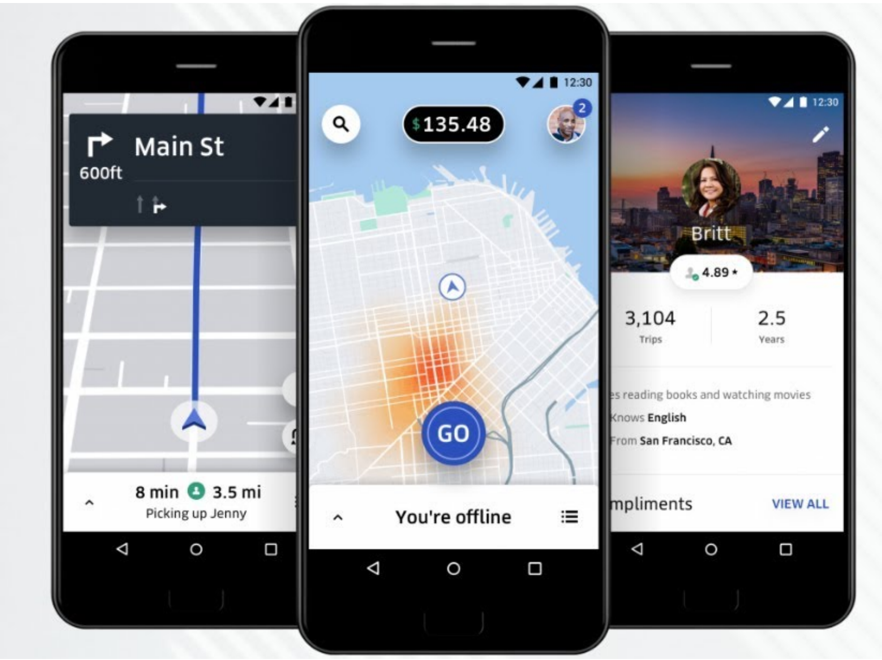
- RL for centrally controlled taxi operation
  - Which zone to go to next?
  - Should it pick up a passenger in the current zone, or move to another zone to maximize its future profits?

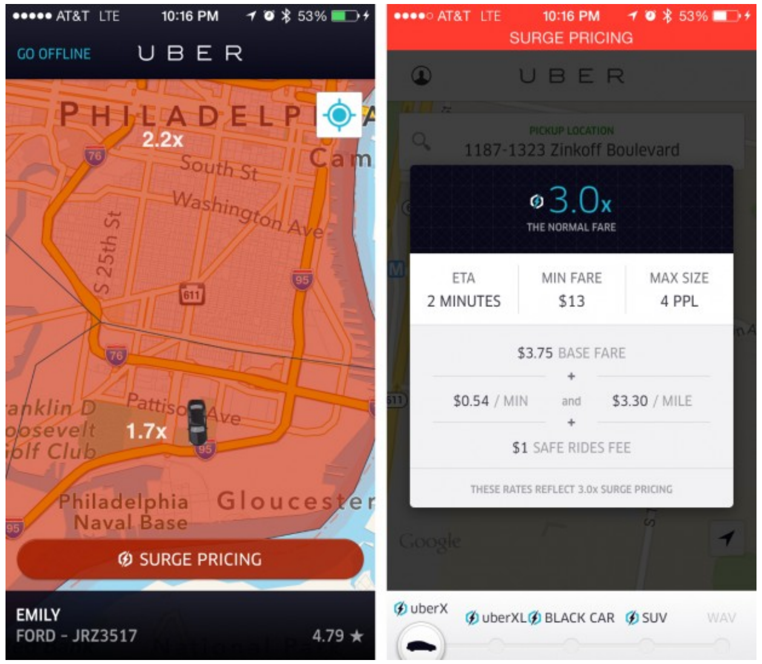


Driver-based Ride hailing systems

- **“Drivers are independent agents. So they are our customers as much as riders are. We cannot tell the drivers where to move. We can suggest where we think they might want to move”**

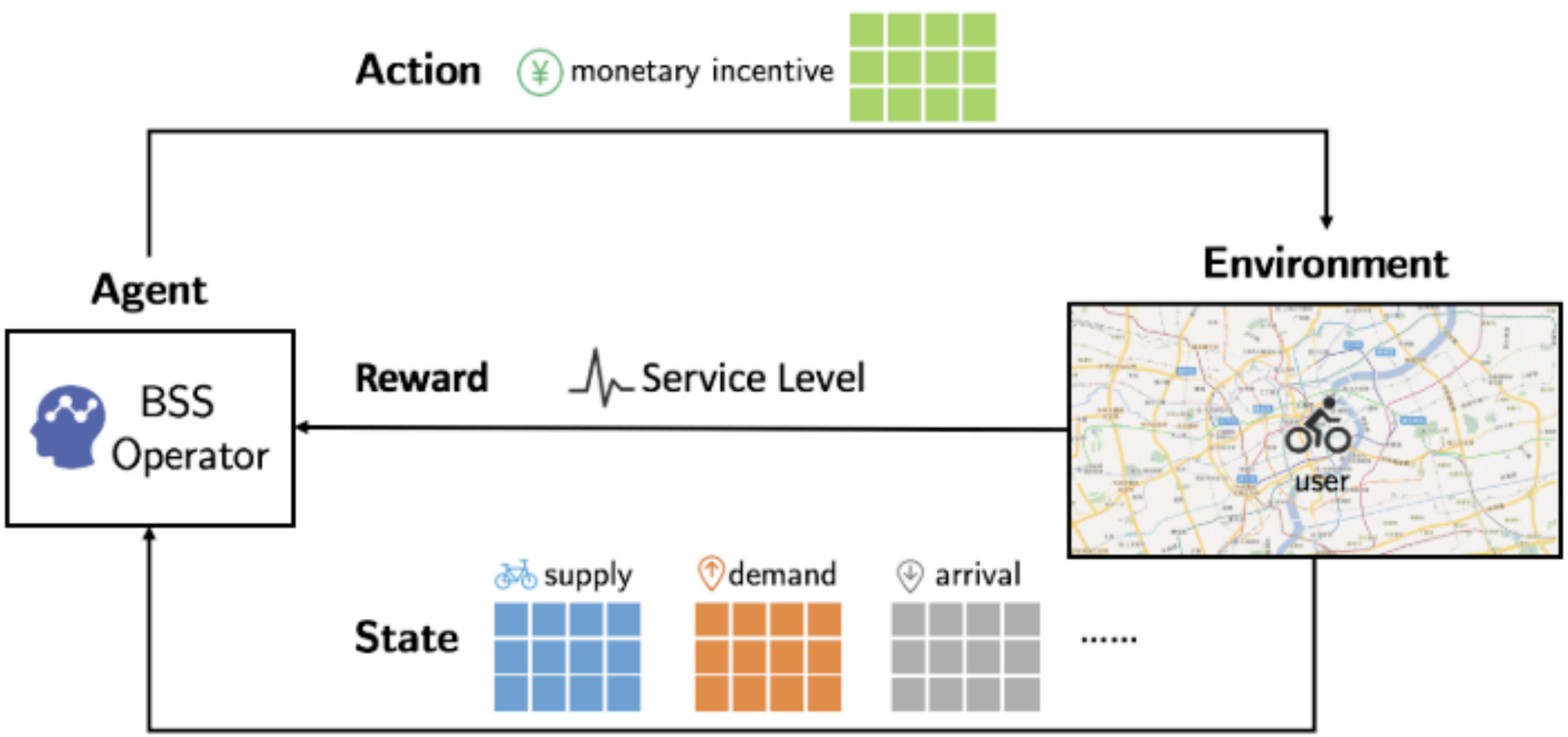
Phillips, R. 2017. Balancing supply and demand in a two-sided marketplace





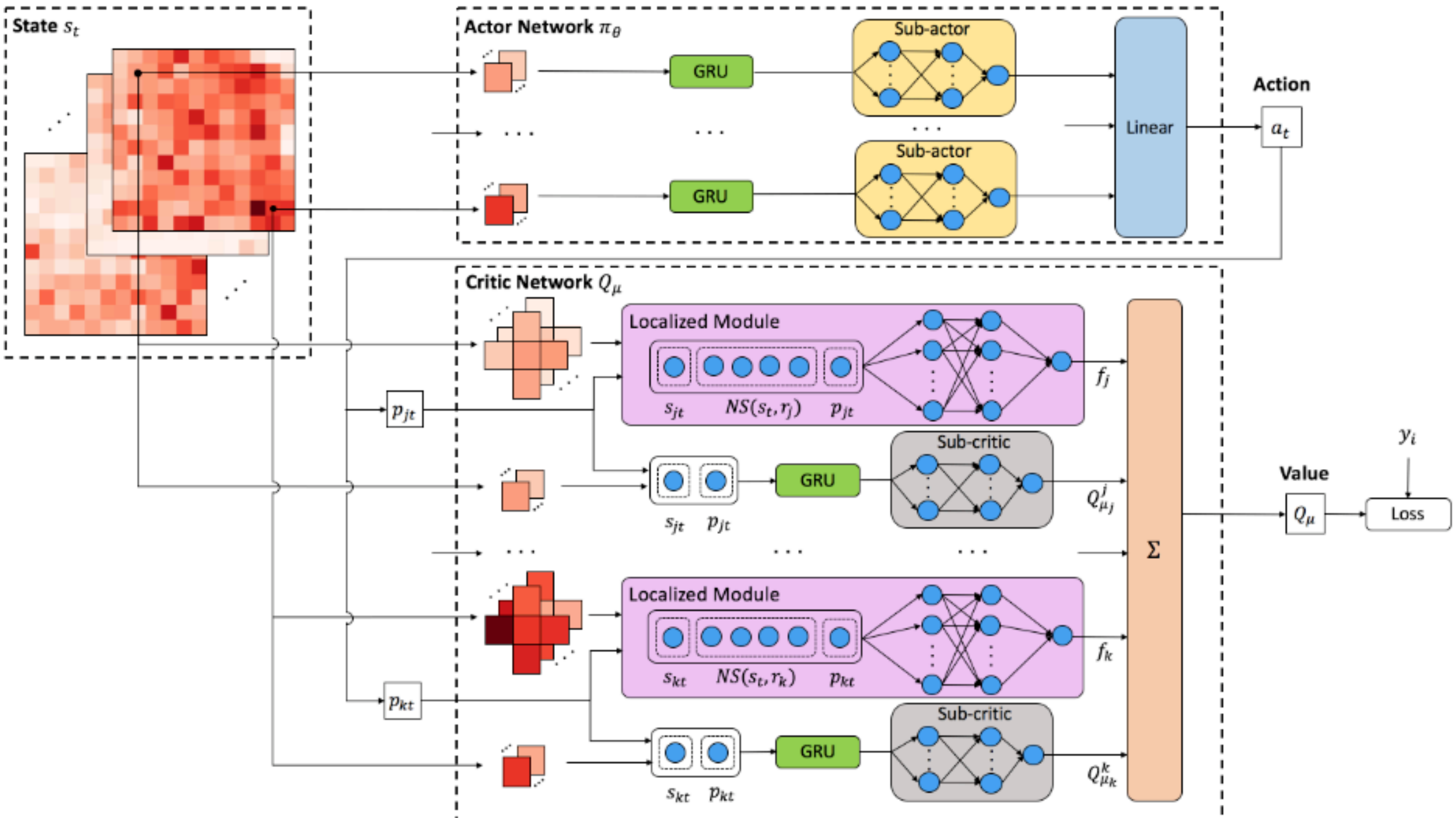
Incentive design

- Incentivize drivers to move to zones that are poorly served



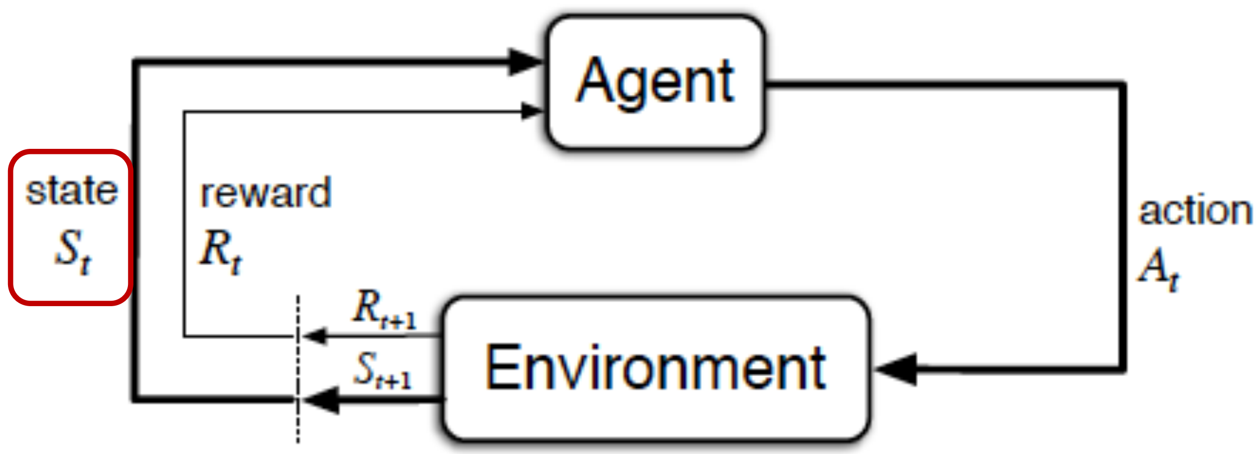
Incentive design

- RL for incentive design



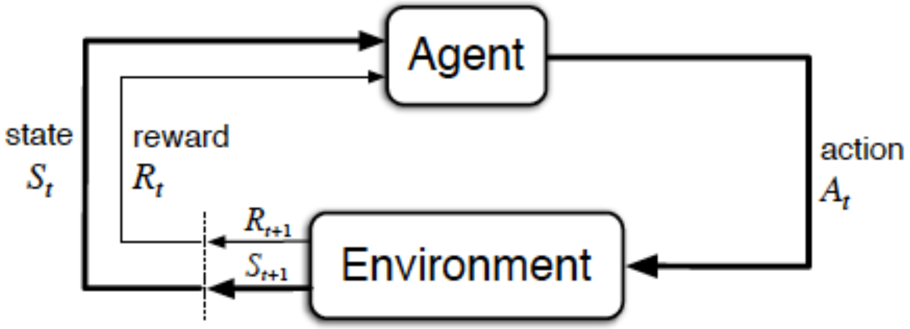
Representation

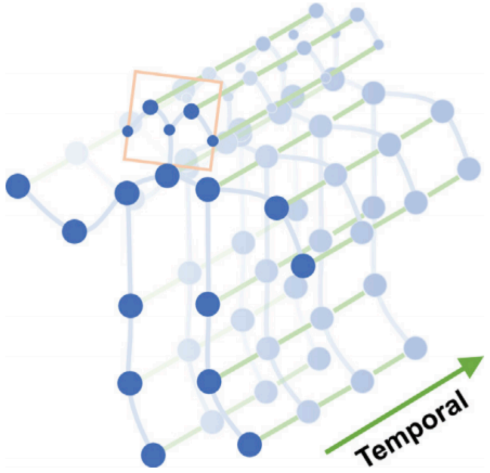
- Agent-environment interaction in a MDP
- How should we represent the state?



Representation

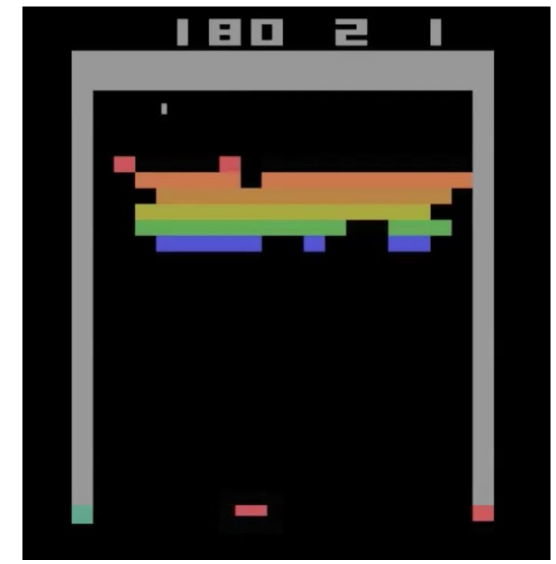
- State representation
  - Atari games
    - Pixels
  - Robotics
    - Joints

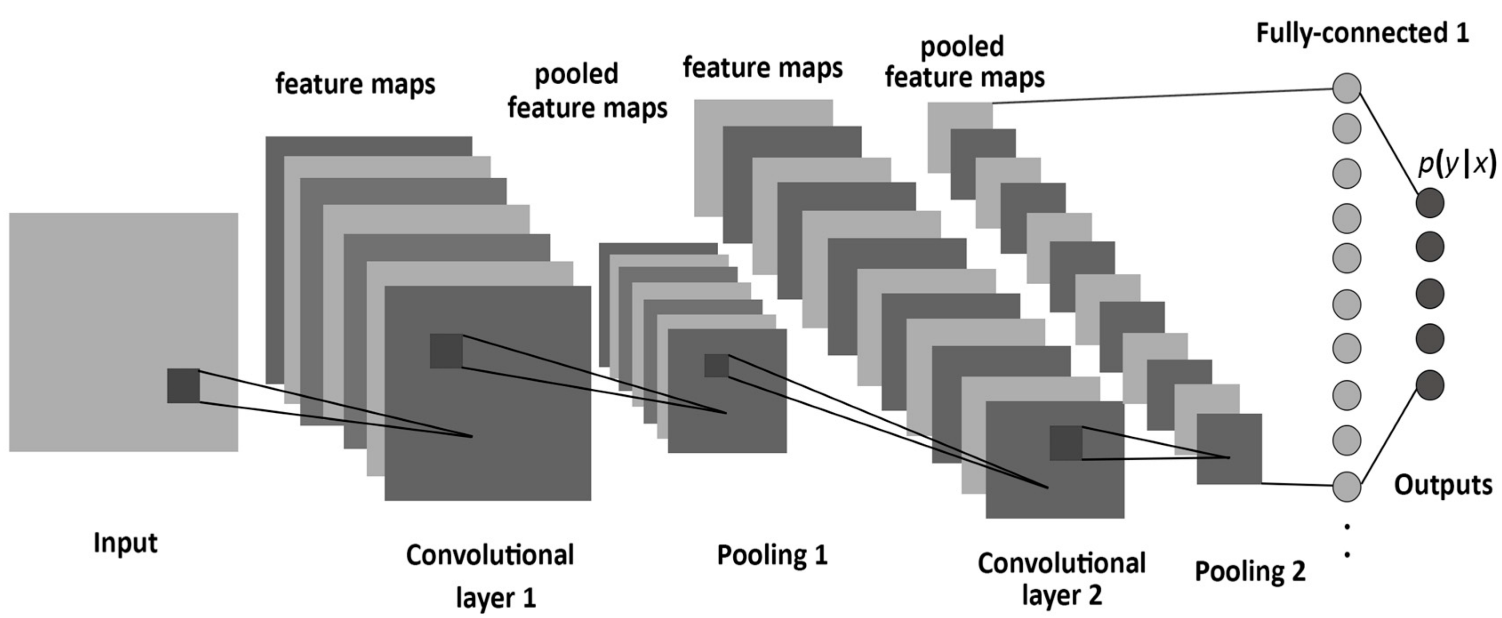




Representation

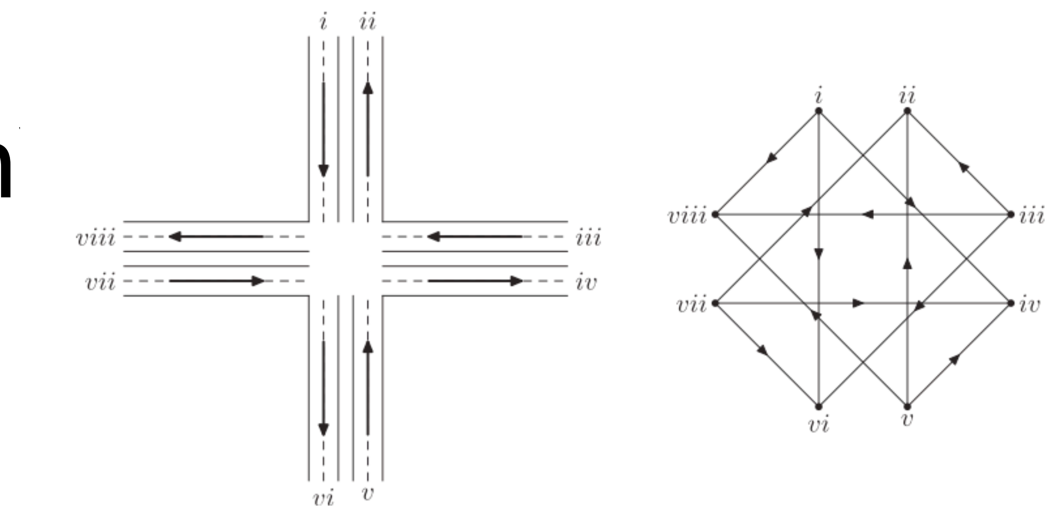
- State representation
  - Atari games
    - Pixels
    - **CNN as feature extraction layer**
  - Robotics
    - Joints



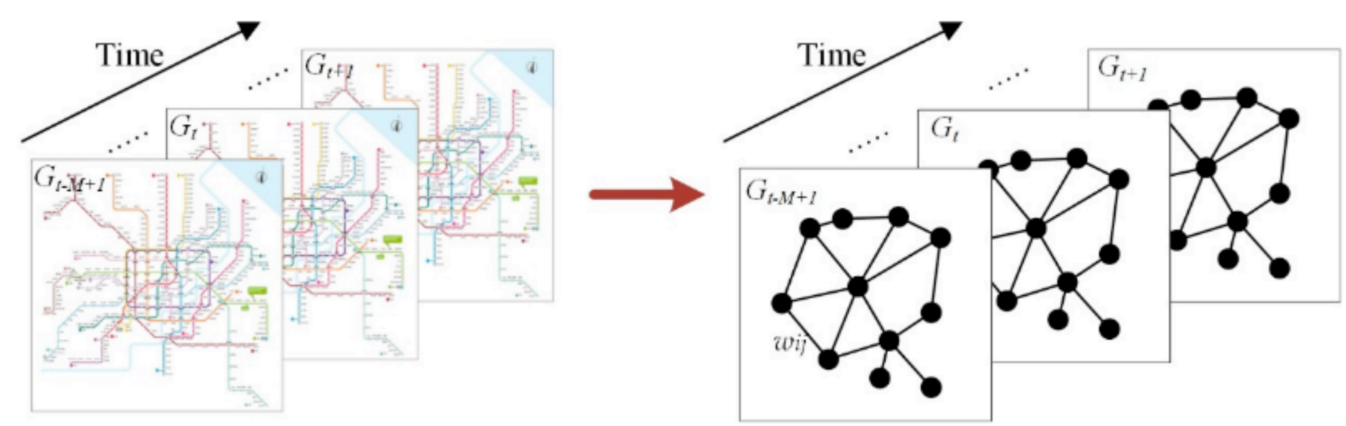


Representation

- What about transportation

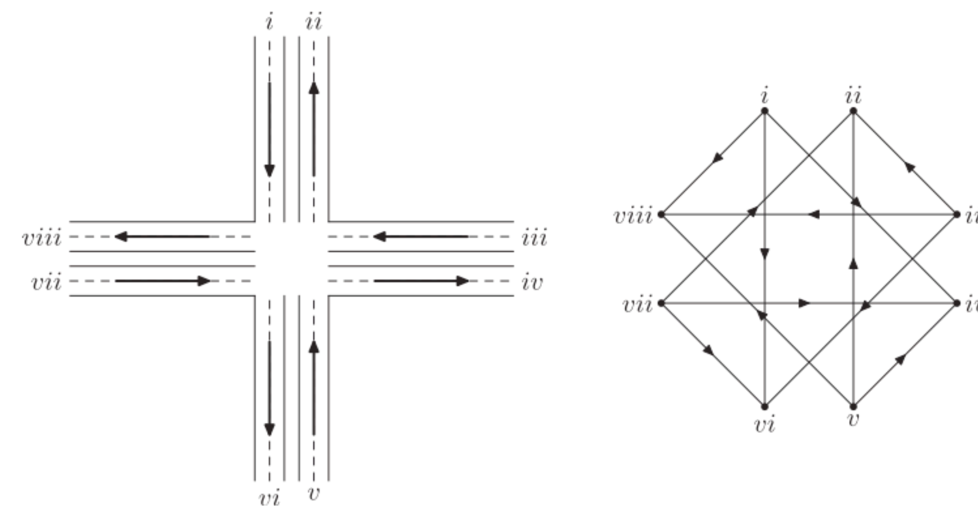


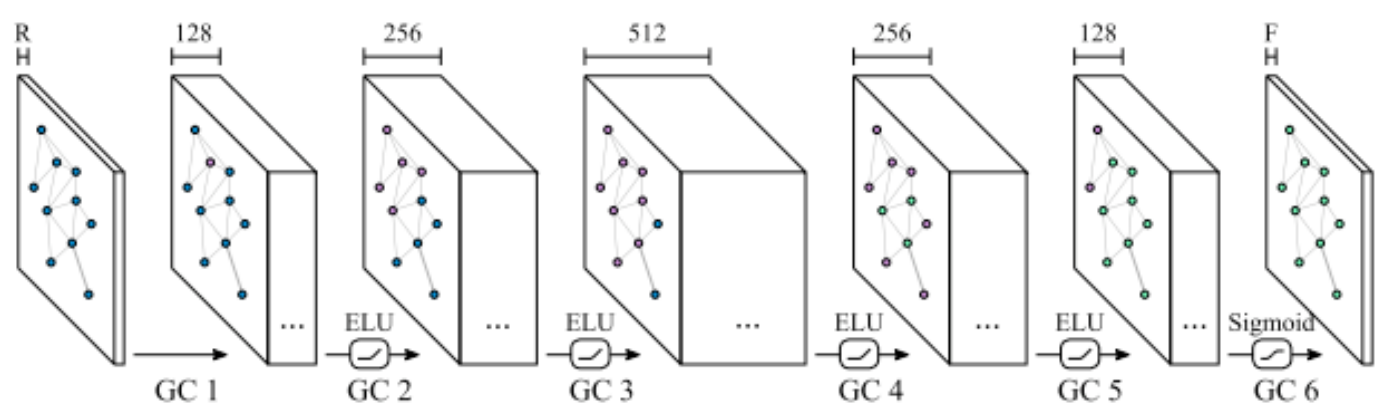
- Typical approach:
  - Graphs!



Representation

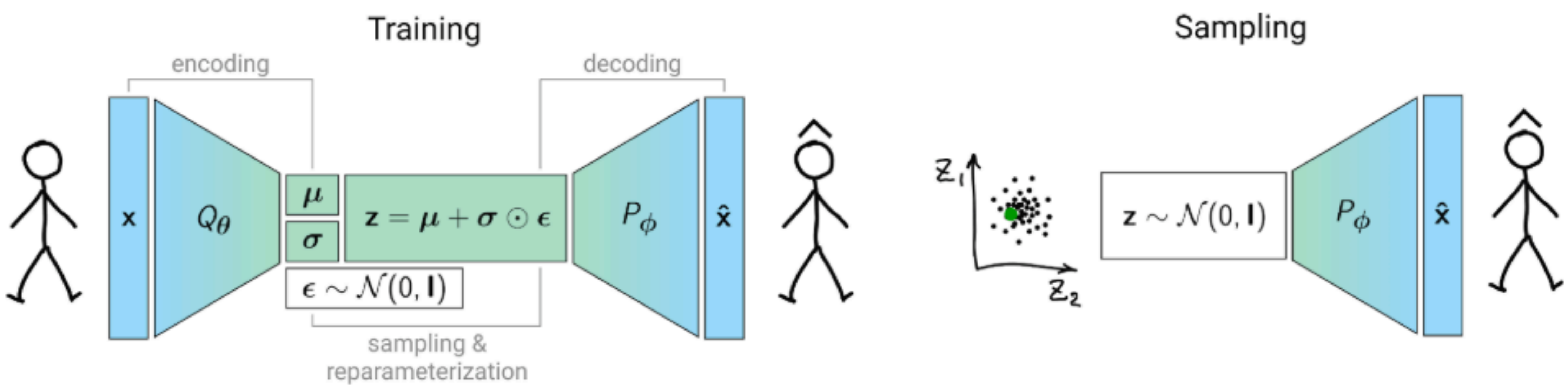
- What about transportation
- Typical approach:
  - Graphs!
- Could we use Graph Convolutional models?





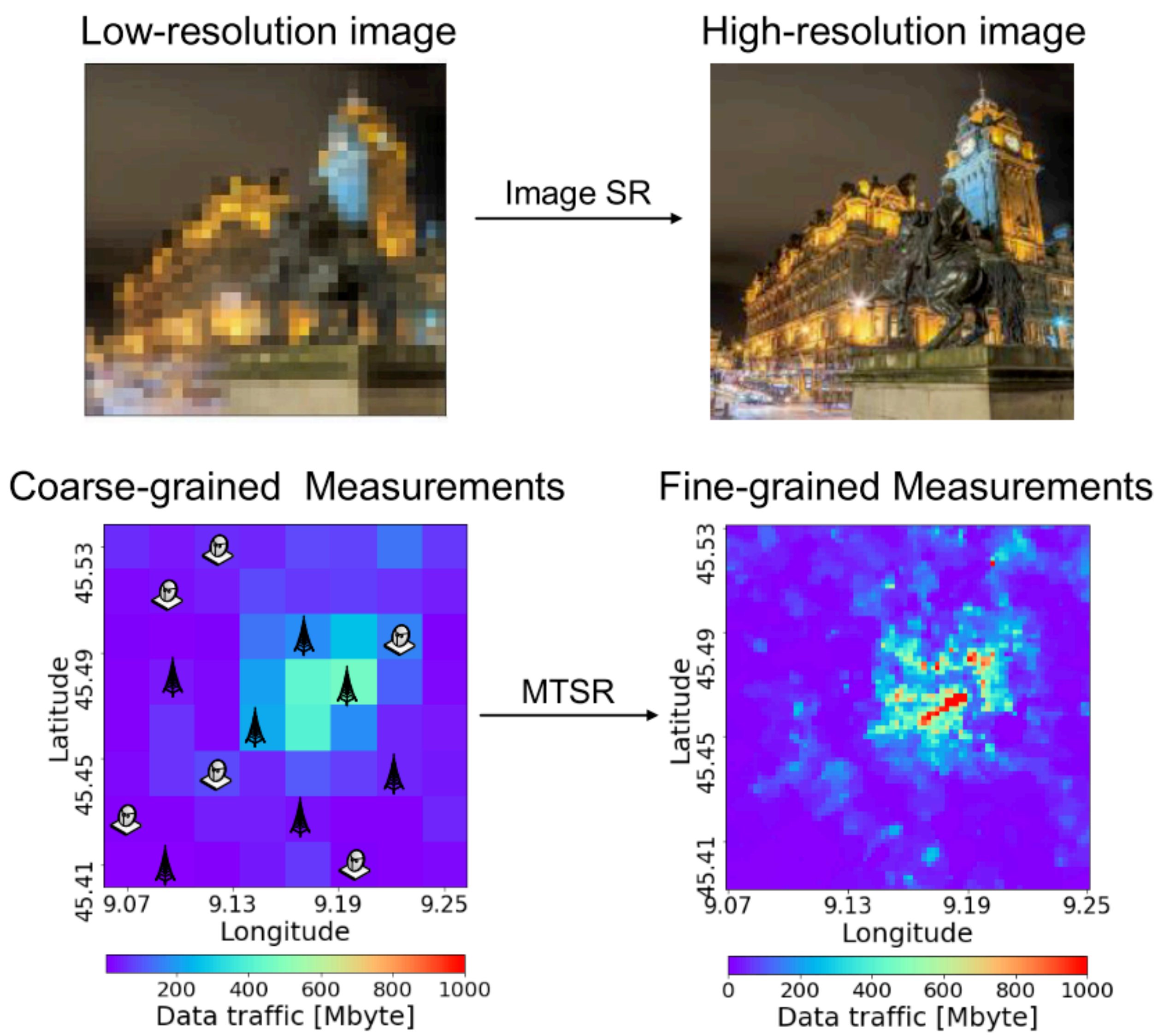
Population synthesis

- How can we learn a realistic representation of the population?



Population synthesis

- How can we learn a realistic representation of the population?



Planning Process

Conventional Travel Forecasting Approach

Data Inputs

Inventories and forecasts of population, land uses, travel behavior, etc.

Trip Generation

Predicts number of trips produced and attracted in a given zone

Trip Distribution

Produces trip production and attraction for each zone

Modal Split

Predicts mode share typically for auto and public transport (can include walk, bike)

Trip Assignment

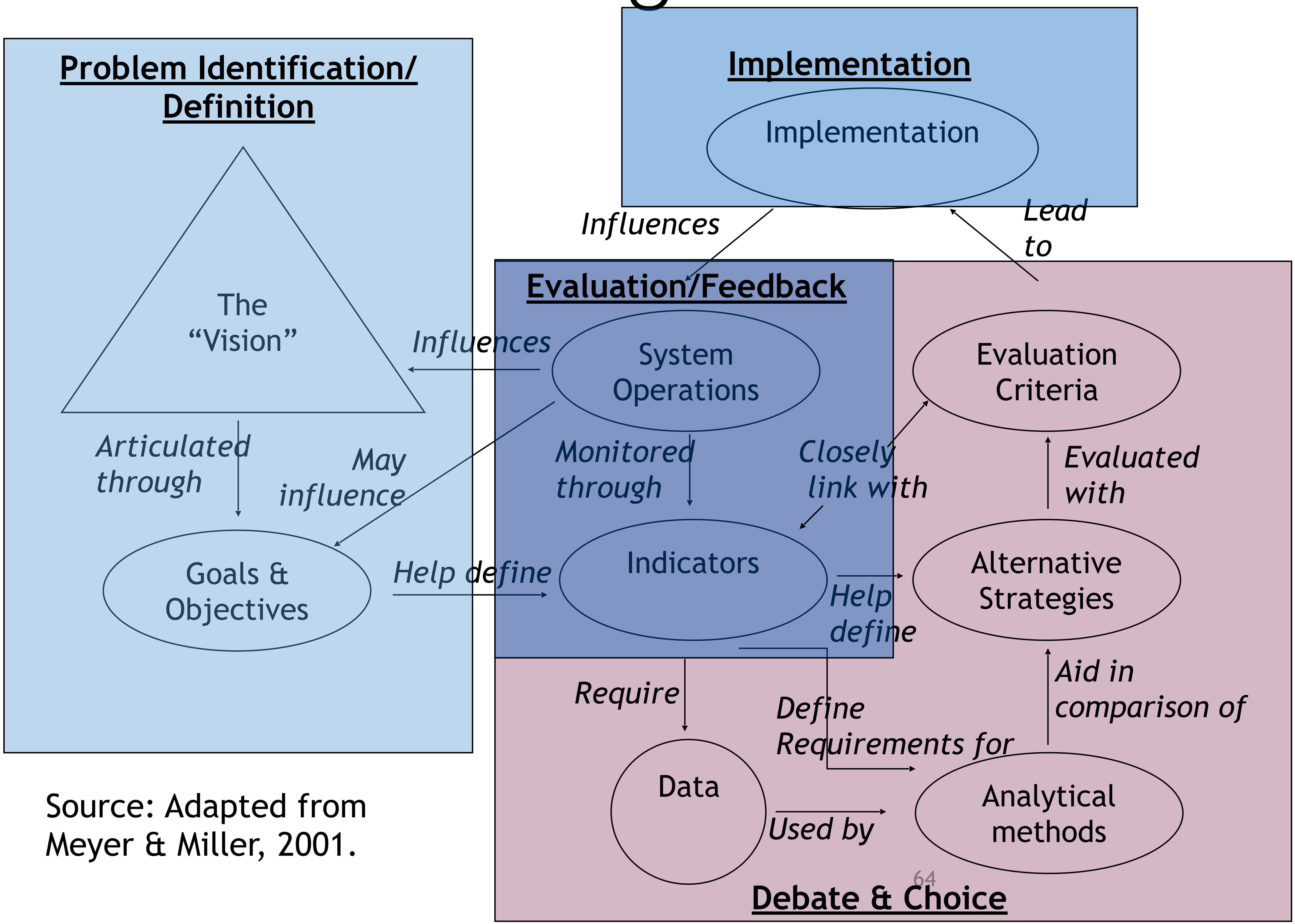
Assigns trips to their respective networks

System Outputs

Provides, for each link, data including traffic volumes, speeds, vehicle mix

Evaluation

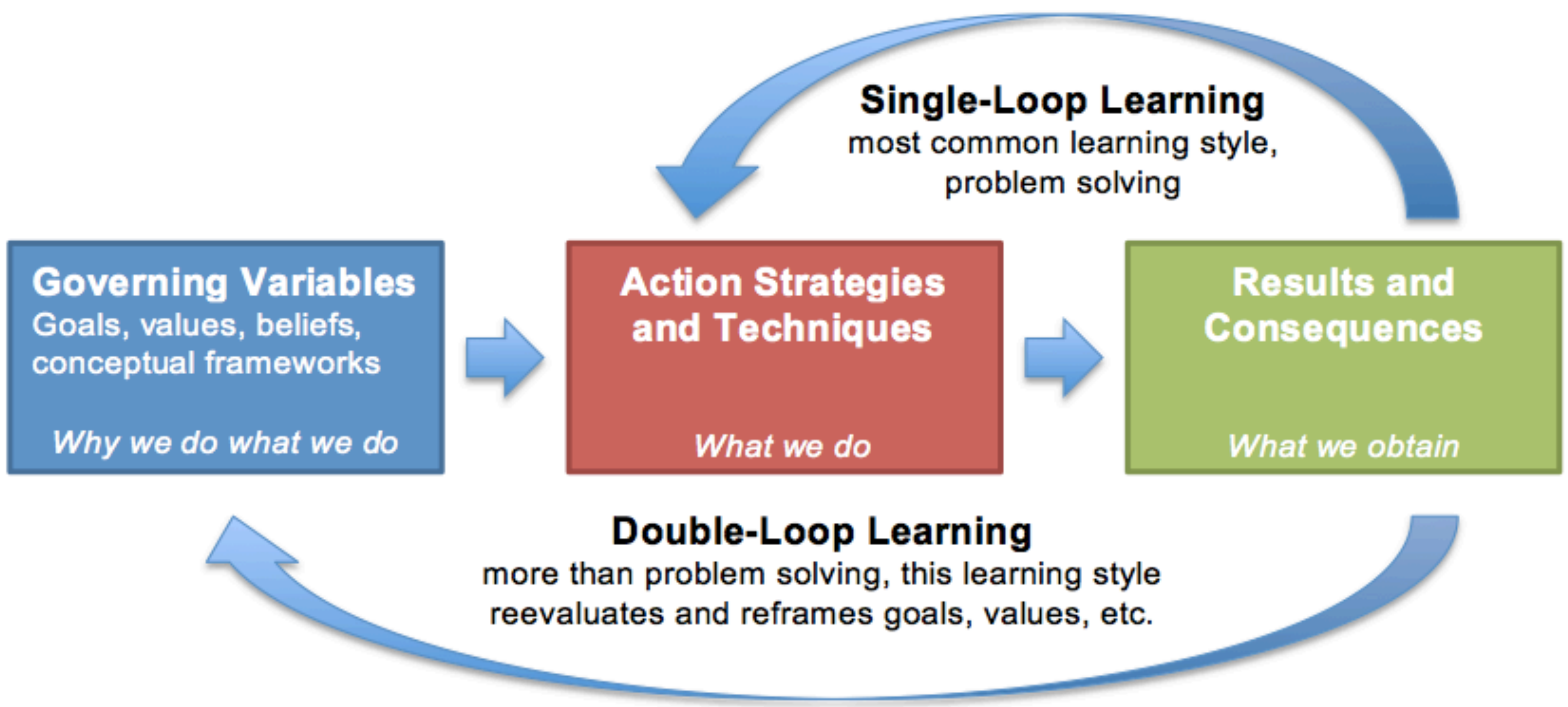
Rational Planning



Source: Adapted from Meyer & Miller, 2001.

Cracks in the rational model….

- Mutual learning: planners + people (Friedmann, 1973)
- Reflective Practitioner (Schon, 1983): reality is a social construction; planner must participate in construction of problem; technical tools fit in reflection in action



Tensions in DNN and Transportation

Tensions in DNN and Transportation

- Role of Model: explain or predict
- Role of Theory: Generic model vs domain specific
- Simple to complex vs complex to simple
- Causality vs correlation
- Understand vs action (control)
- Knowledge production: discover vs create

Criteria of a good model

- Prediction accuracy
- Interpretability (FAT)
- Robustness (security)
- Sparsity
- Practicality
- Difficulty

DNN and Behavioral Theory

Deep Neural Networks for Choice Analysis

Shenhao Wang

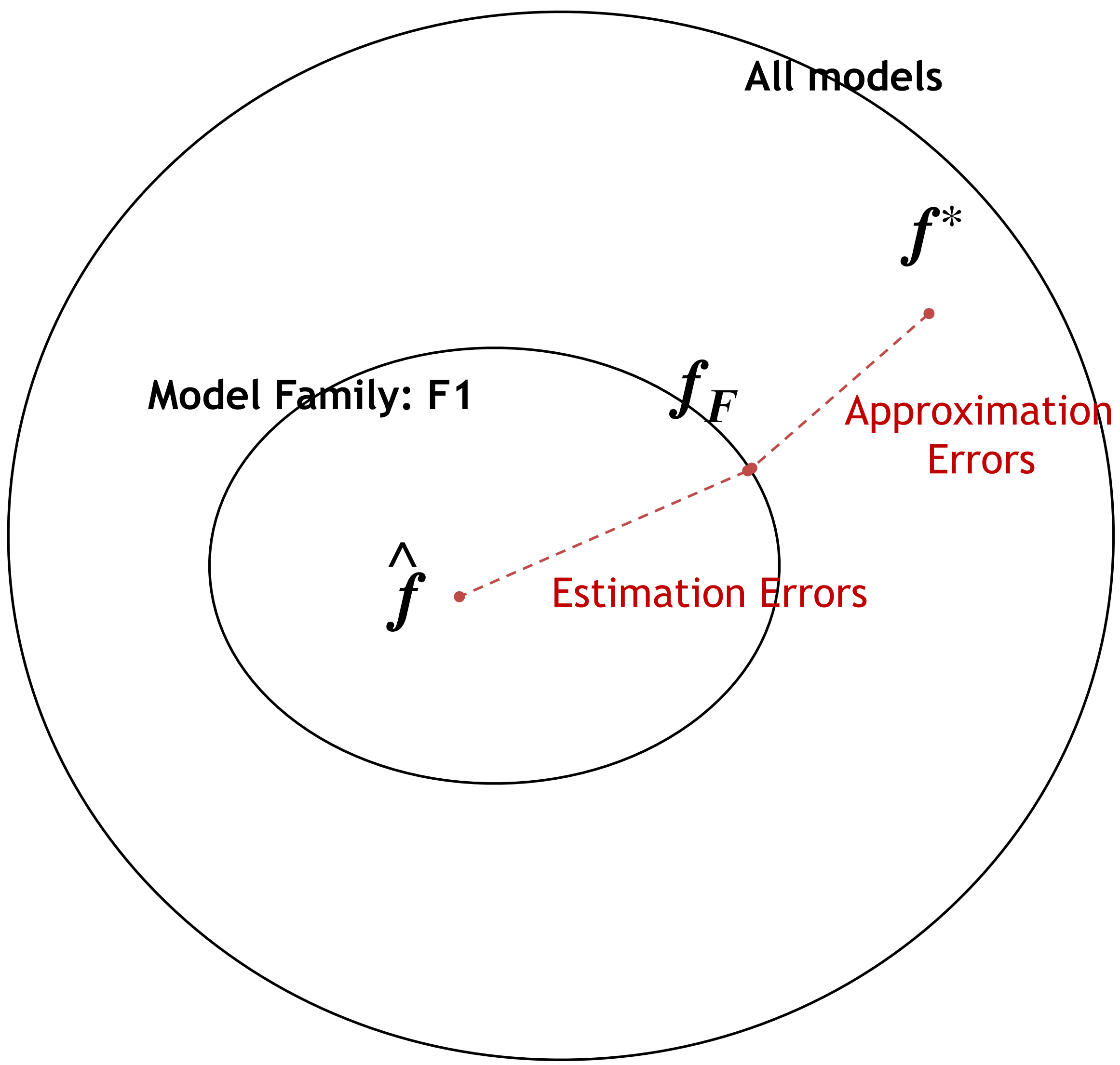
Dissertation defense for the degree of
Doctor of Philosophy in Computer and Urban Science
September, 2019

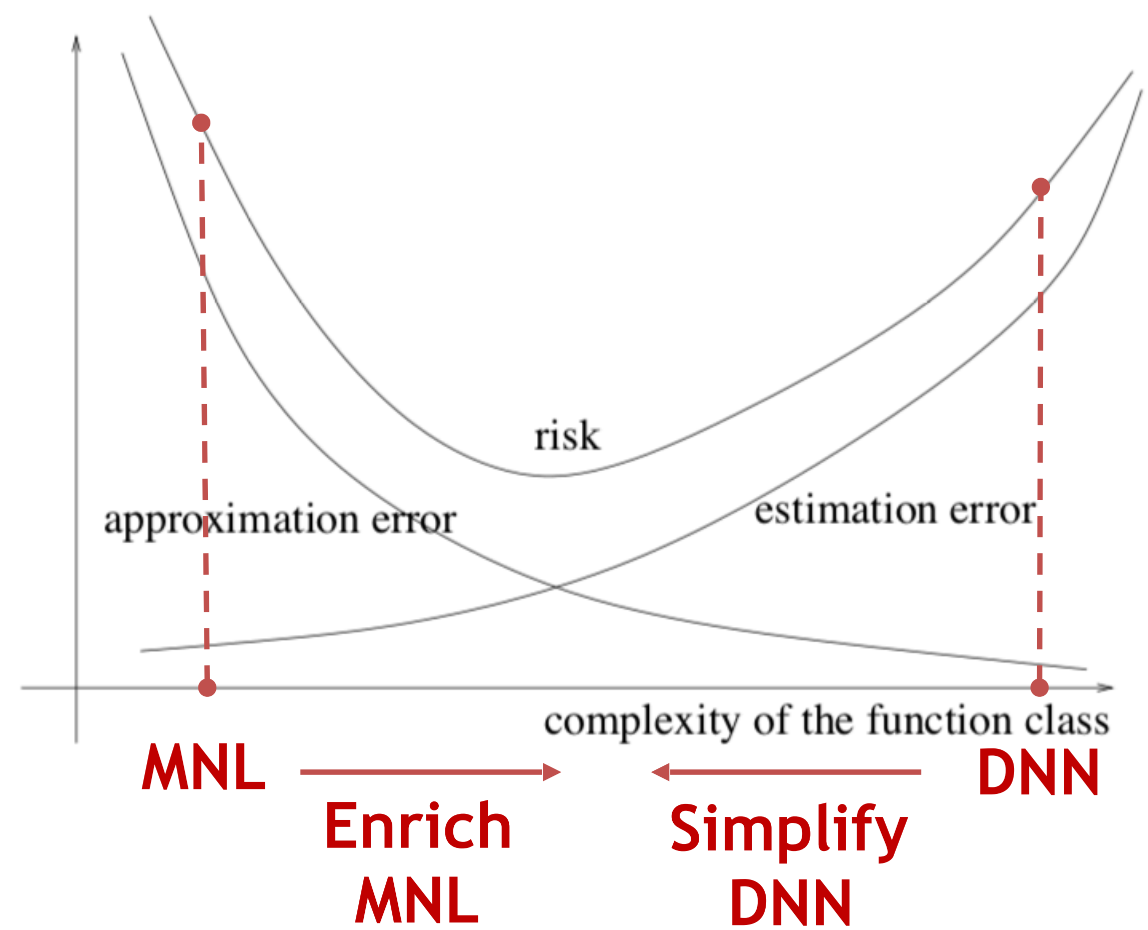
Two Modeling Paradigms for Choice Analysis

Deep Neural Networks	Dimension	**Discrete Choice Models**
High	Prediction Accuracy	Low
Automatic	Representation Learning	Handcrafted
Low/Unclear	Interpretability	High
Difficult/Flexible	Architectural Design & Regularizations	Unclear
Low	Robustness	High

Prediction Accuracy: Nijkamp et al., (1996); Xie et al., (2003); Cantarella & de Luca, (2005); Celikoglu (2006); etc.
Representation Learning: LeCun et al. (2015); Bengio et al. (2013); etc.
Interpretability: Kim and Doshi-Velez (2017); Lipton (2016); Montavon et al. (2018); Ribeiro et al., (2016); etc.
Architectural Design & Regularizations: Krizhevsky, et al. 2012; Zoph et al., 2017; Martin and Bartlett (2009); etc.
Robustness: Szegedy (2014); Goodfellow (2015); Papernot (2016); Kurakin (2017); Matthew and Jegelka (2017), etc.

Prediction Errors = Approximation + Estimation Errors

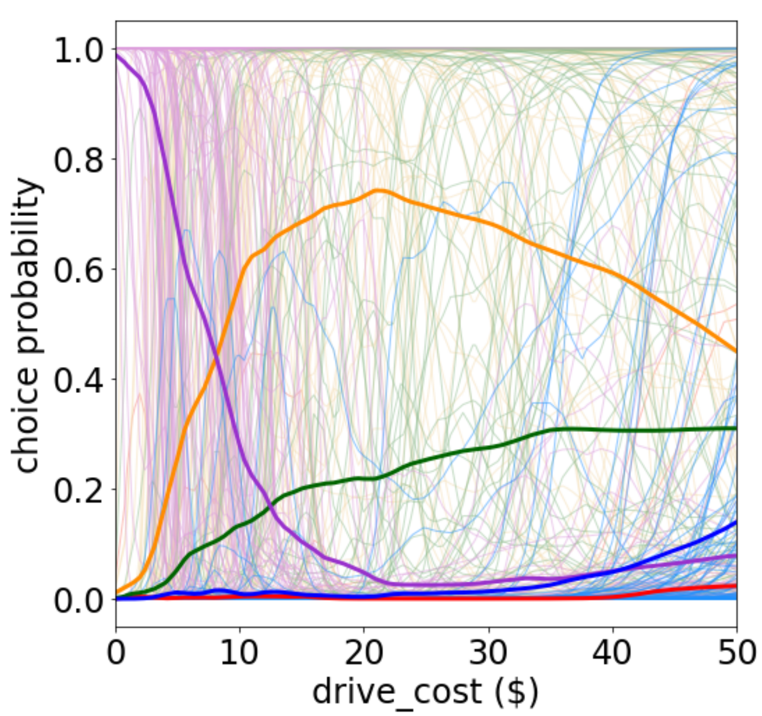




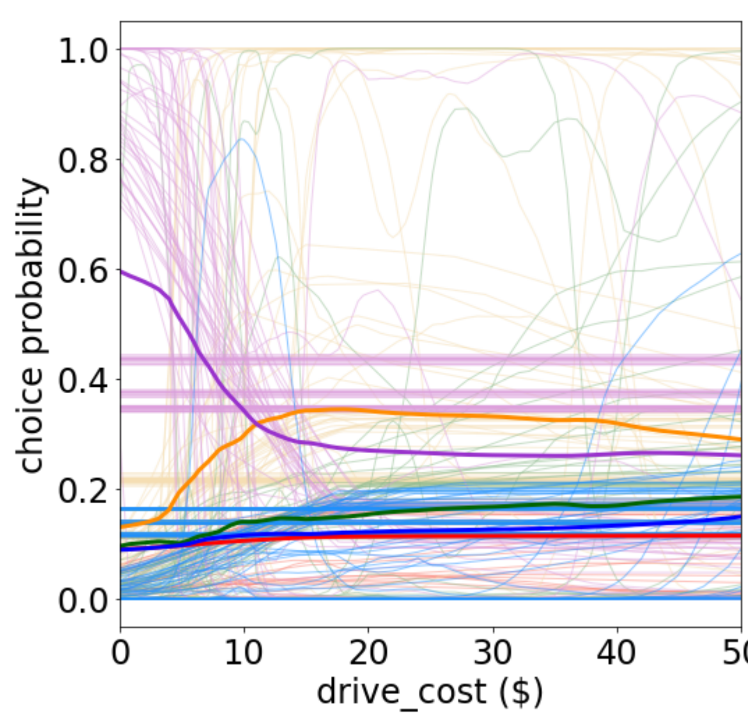
Substitution Patterns of Five Alternatives

$$s_{k_1}(x_j; x_{\backslash j})/s_{k_2}(x_j; x_{\backslash j})$$

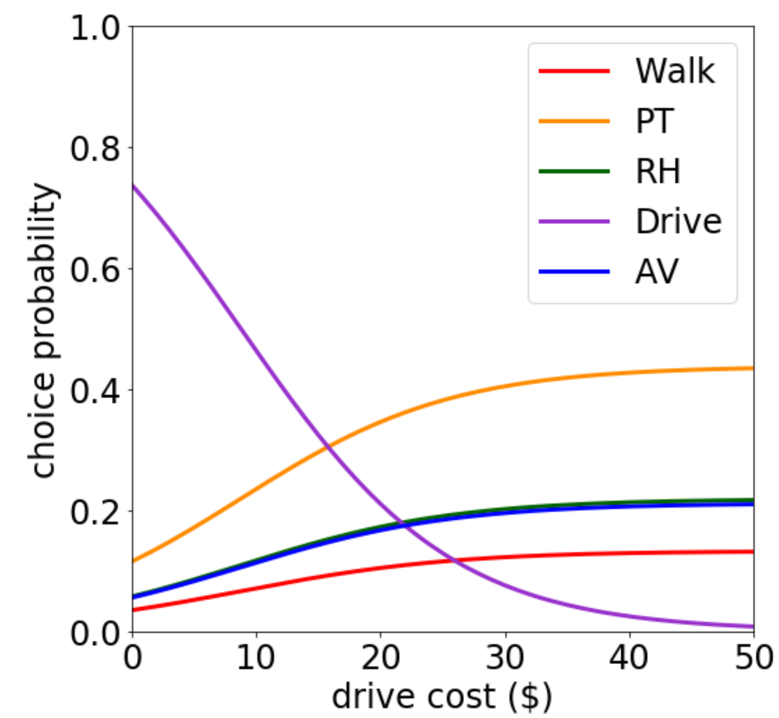
5L-DNNs



HP-DNNs



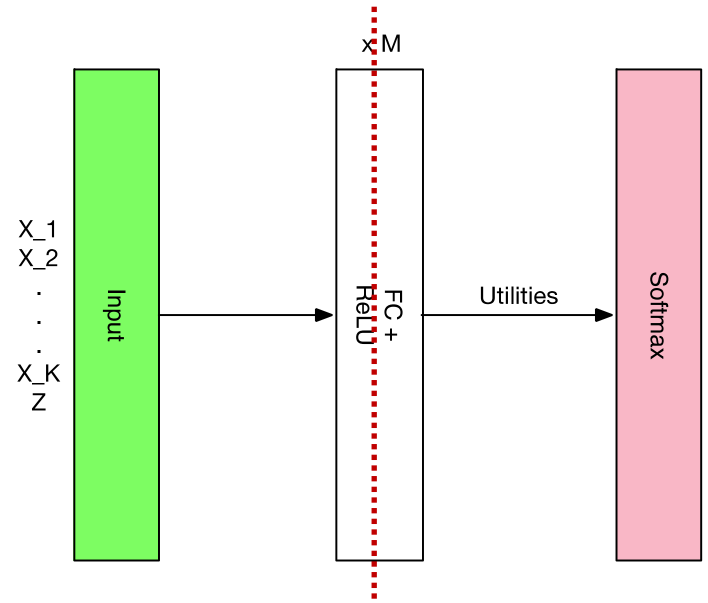
MNL



Synergy between DNN and DCM

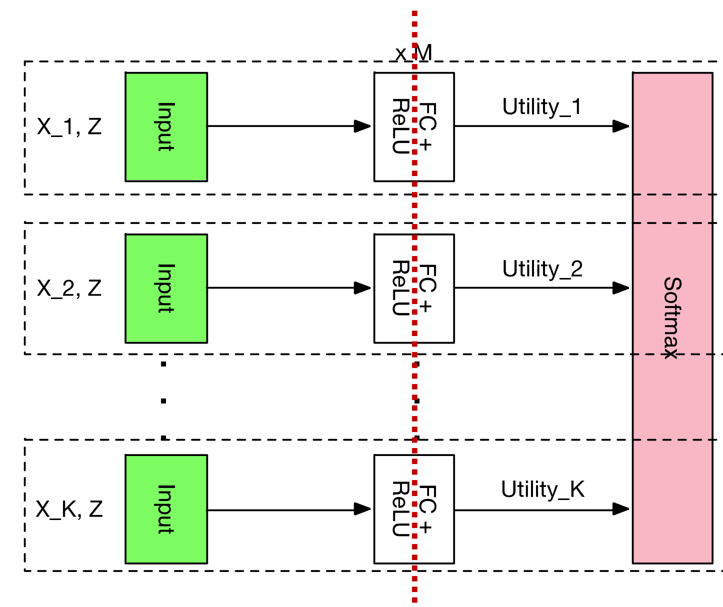
Using Utility Theory to Design DNN

Fully connected DNN (F-DNN)



$$V_k = f(x_1, x_2, ...x_K, z)$$

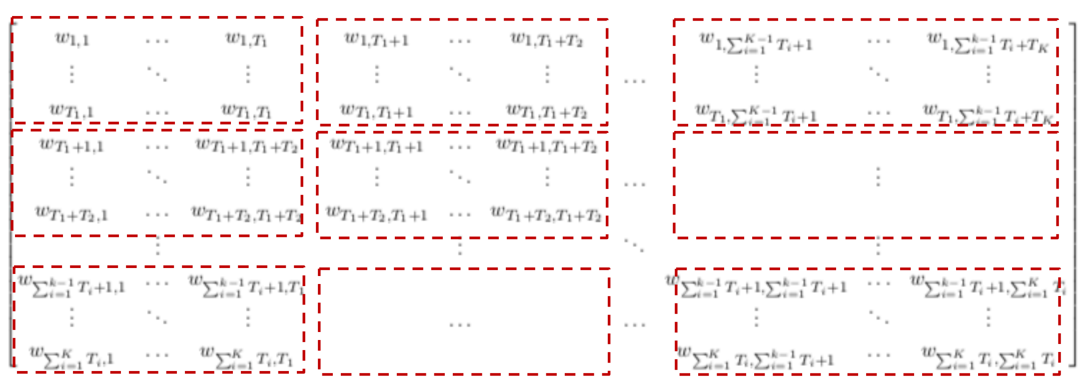
DNN with alternative-specific connection (ASU-DNN)



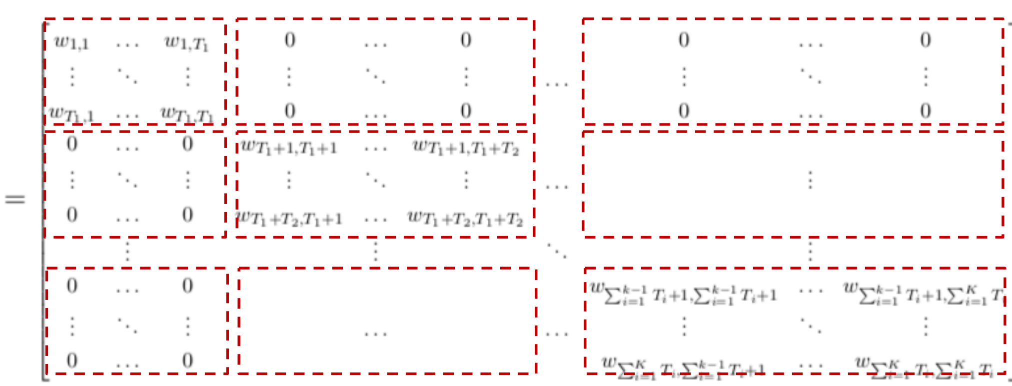
$$V_k = f(x_k, z)$$

Sparsity as Regularization Introduced by Utility Theory

$W_{F-DNN} =$



$W_{ASU-DNN} =$



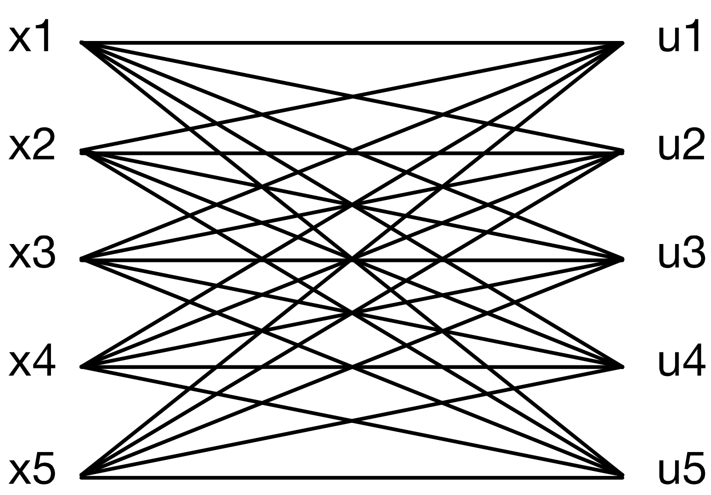
Reduce Estimation Error

$$\hat{R}_n(F_2|_S) \leq \frac{(\sqrt{2\log(D)}+1)\sqrt{\frac{1}{N}\sum_{i=1}^{N}||x_i||^2}}{\sqrt{N}} \times c^D T^D$$

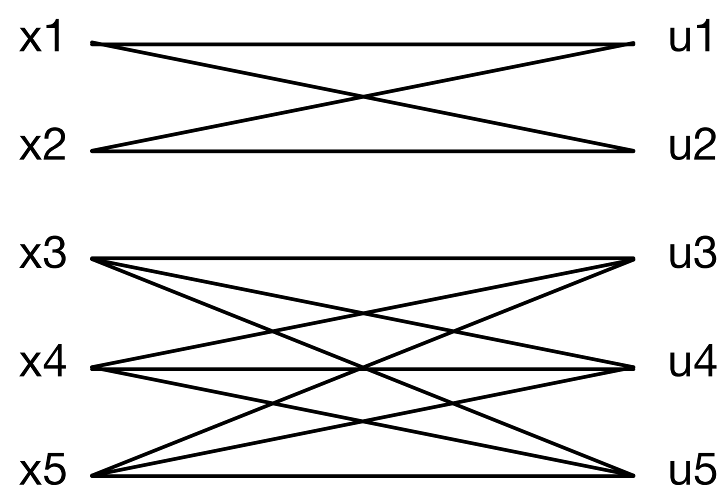
$$\hat{R}_n(F_1|_S) \leq \frac{(\sqrt{2\log(D)}+1)\sqrt{\frac{1}{N}\sum_{i=1}^{N}||x_i||^2}}{\sqrt{N}} \times \frac{c^D T^D}{K^{D/2}}$$

DNN Design with Utility Connectivity Graph

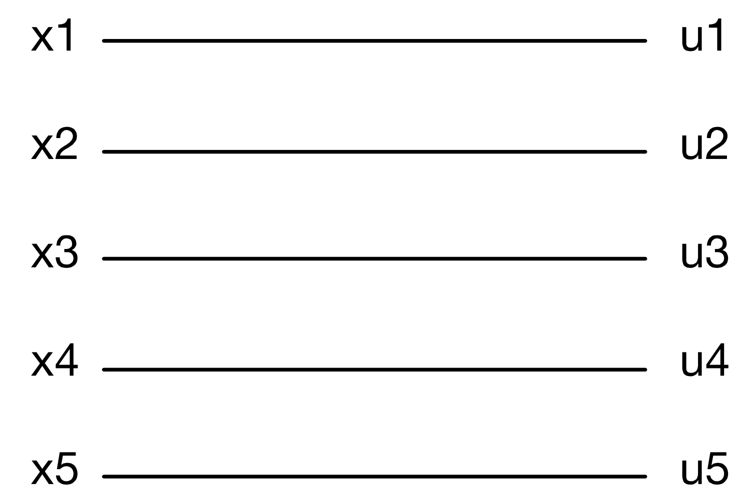
F-DNN



NSU-DNN



ASU-DNN



$$A_{F-DNN} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

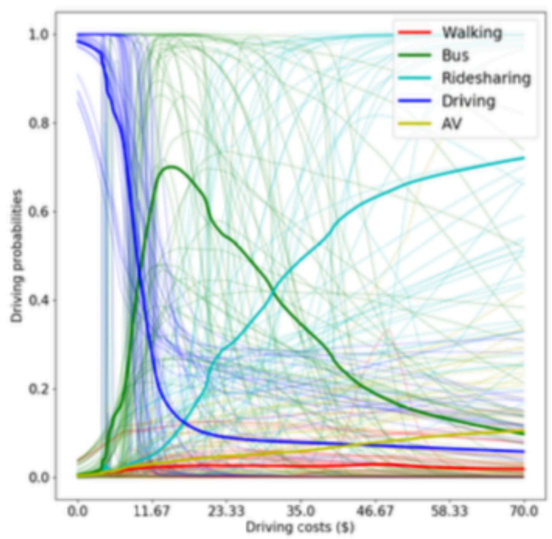
$$A_{NSU-DNN} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$A_{ASU-DNN} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

1. **Implementation in DNN: Group-LASSO;**
2. **Local vs. Global Connectivity Control**

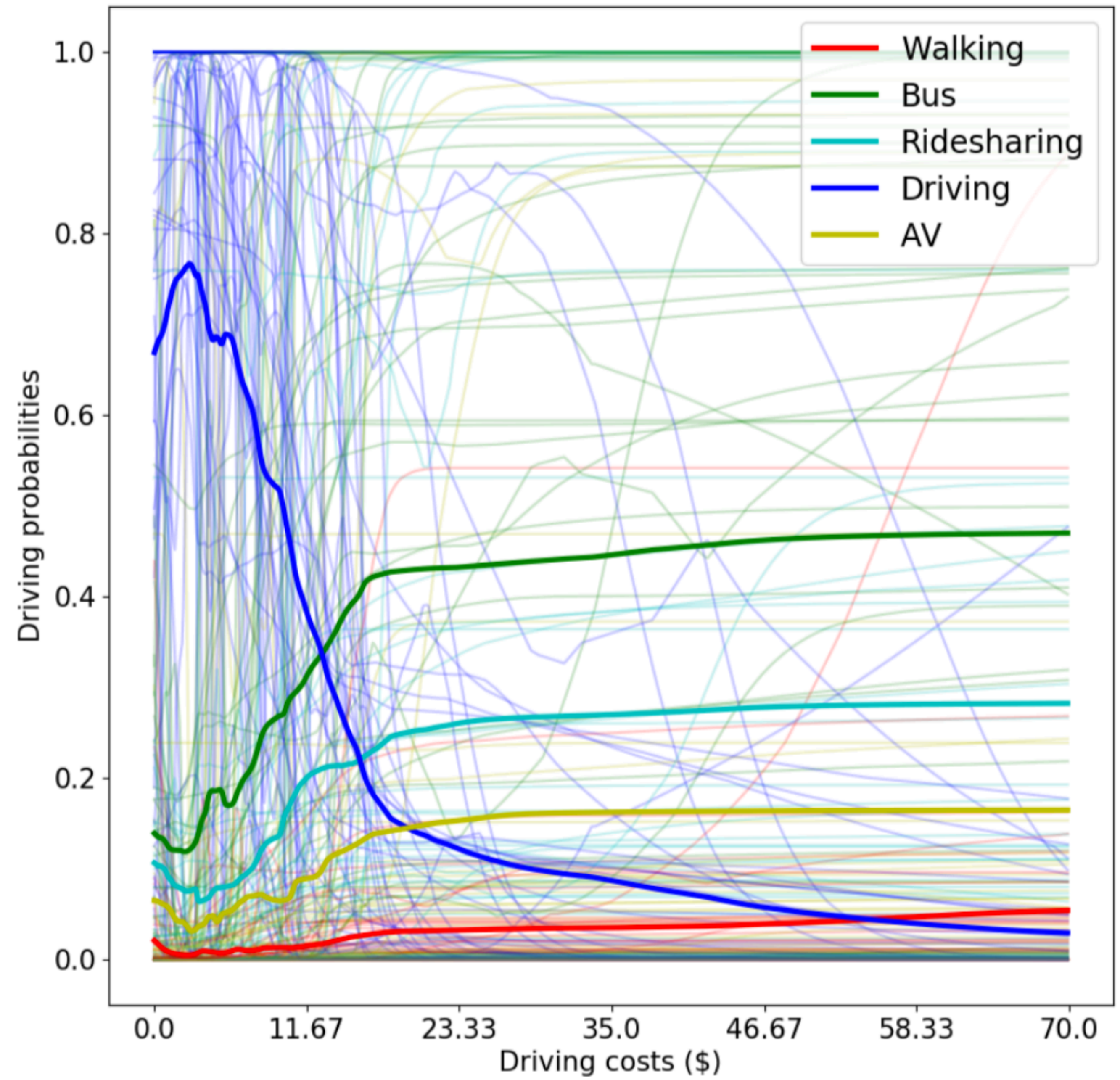
Empirical Result 3: Interpreting the Substitution Pattern

F-DNNs

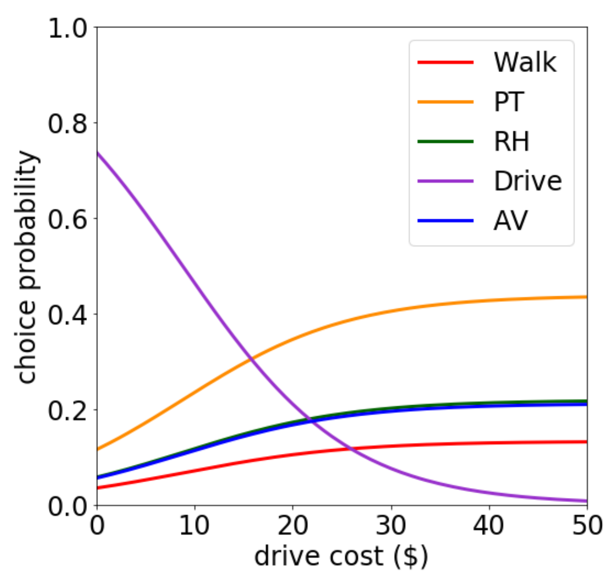


(c) F-DNN (Top 10 Models)

ASU-DNNs



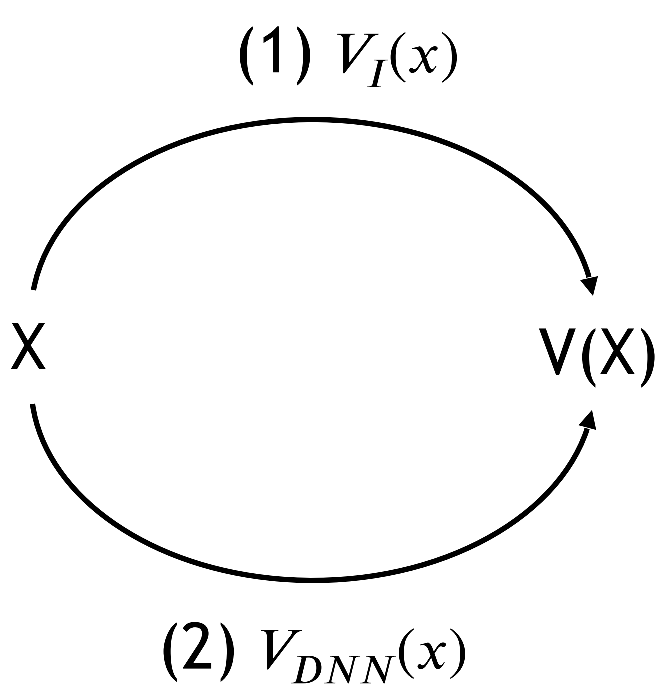
MNLs



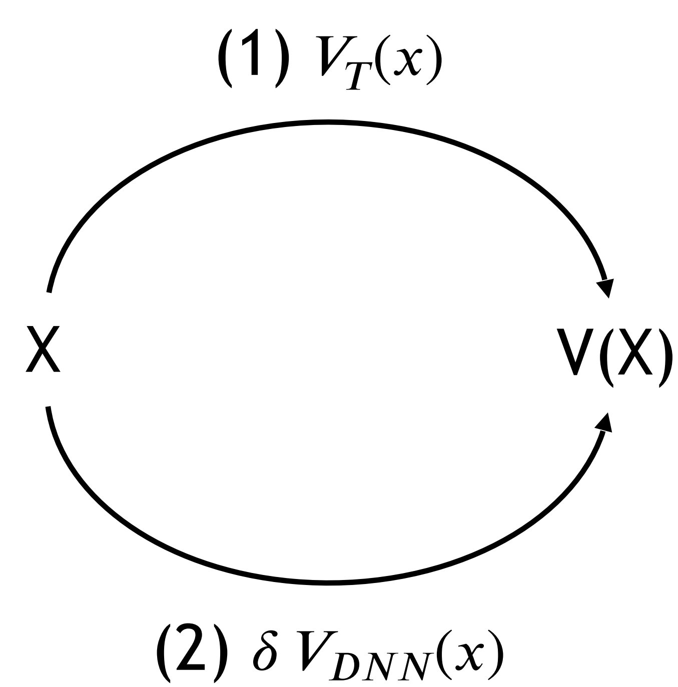
Theory-Based Residual Neural Network (TB-ResNet)

$$V(x) = V_T(x) + \delta V_{DNN}(x)$$

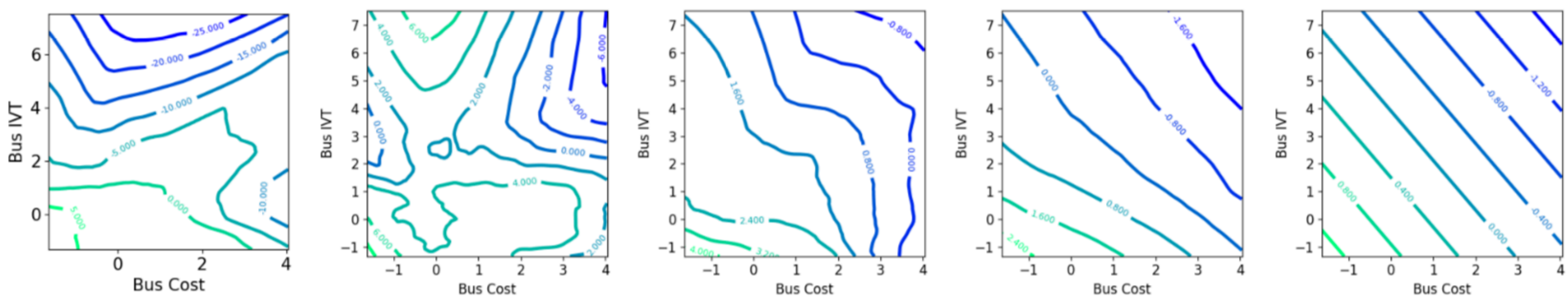
ResNet



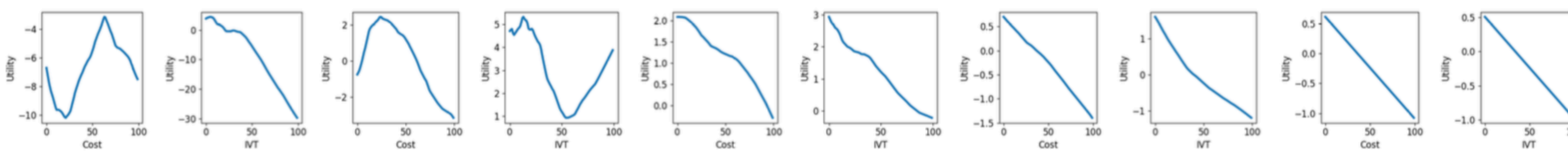
TB-ResNet



2. Interpretability of Utility Function in the CM Scenario

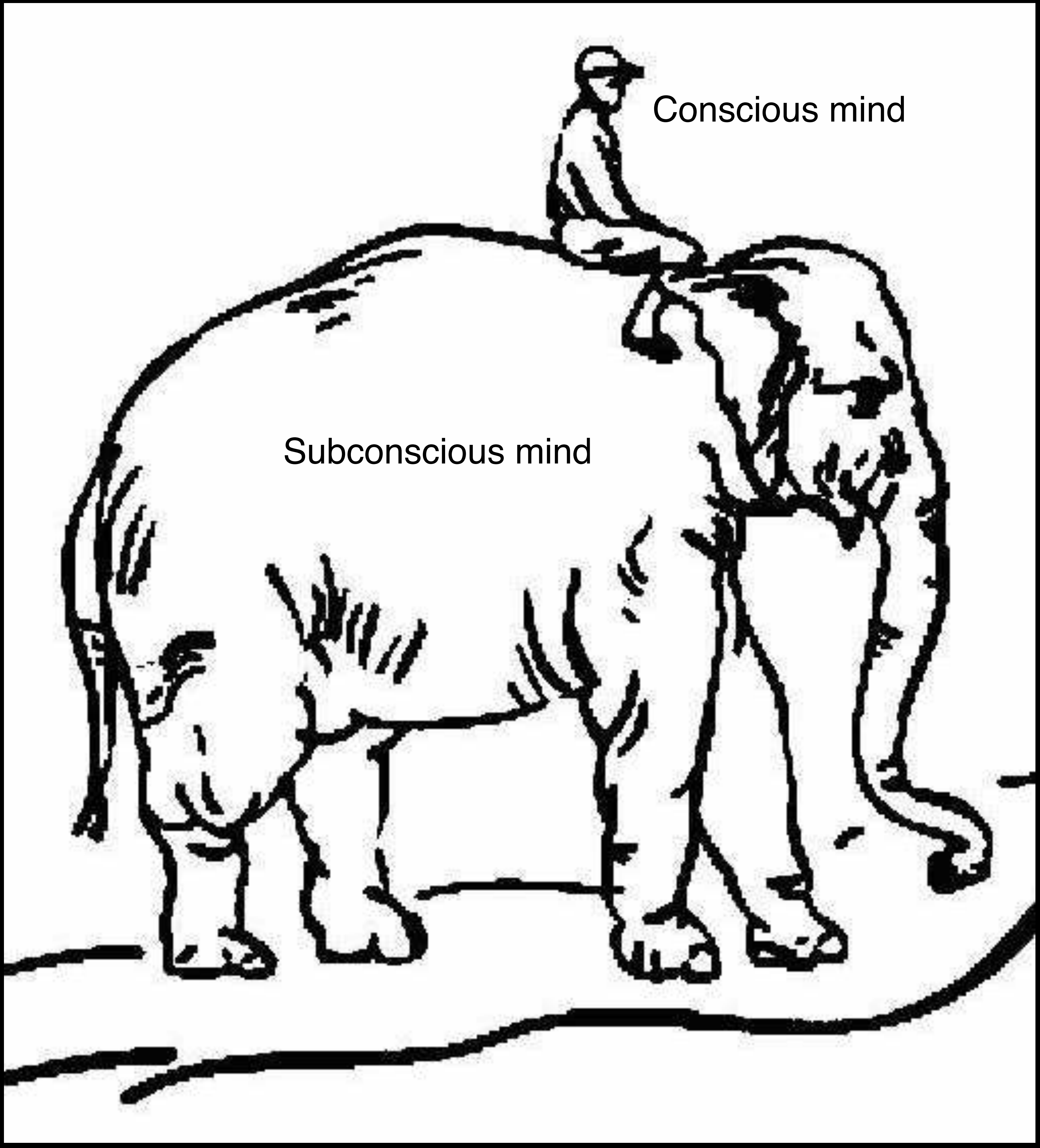


(a) DNN (55.2%) (b) CM Resnet ($\lambda = 1e-10$; 56.4%) (c) CM Resnet ($\lambda = 0.005$; 57.3%) (d) CM Resnet ($\lambda = 0.01$; 56.8%) (e) CM (44.7%)



(f) x0 (g) x1 (h) x0 (i) x1 (j) x0 (k) x1 (l) x0 (m) x1 (n) x0 (o) x1

A Behavioral Perspective


Conscious mind
Subconscious mind

Homo Econ vs. Homo Sapiens

- Rational Agent Assumption
  - perception rationality
  - preference rationality
  - process rationality
- Behavioral deviation
  - loss aversion
  - overweighing small probability
  - mental accounting
  - hyberbolic discounting
  - framing effects
  - altruism
  - price saliency

Behavioral Science

~

Transportation Technology

- Emotional
- Social
- Perceptional

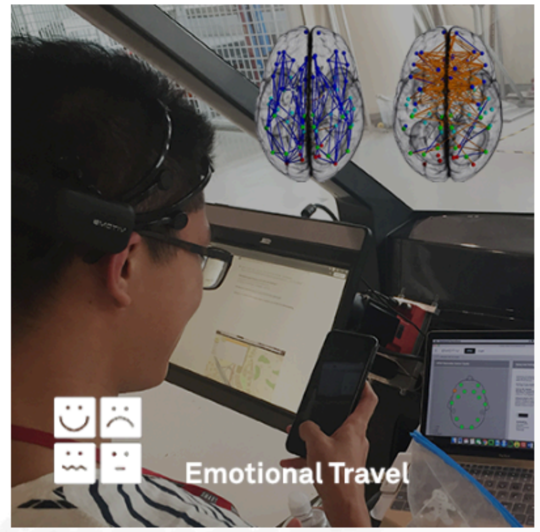
- Sharing
- Automation
- On-demand

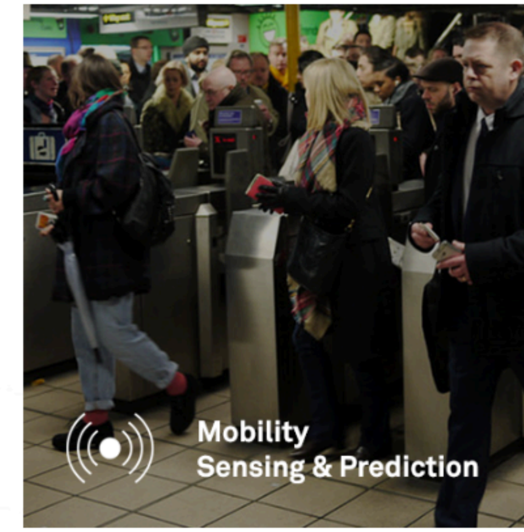
Data Analytics and Machine Learning

JTL Urban Mobility Lab

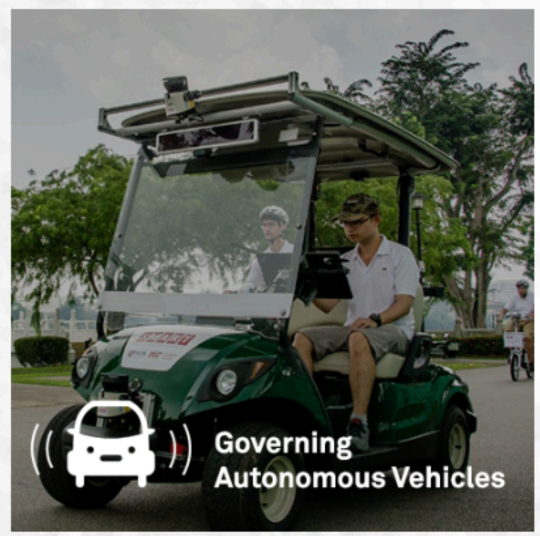
mobility.mit.edu

SECTION I. BEHAVIORAL FOUNDATION OF URBAN MOBILITY

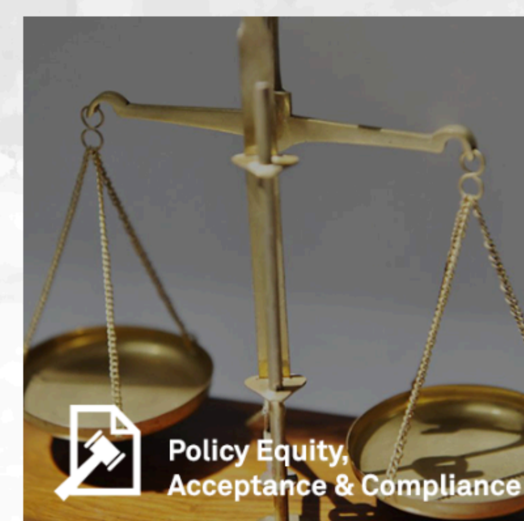




SECTION II. MOBILITY SYSTEM: DESIGN WITH BEHAVIORAL PERSPECTIVE

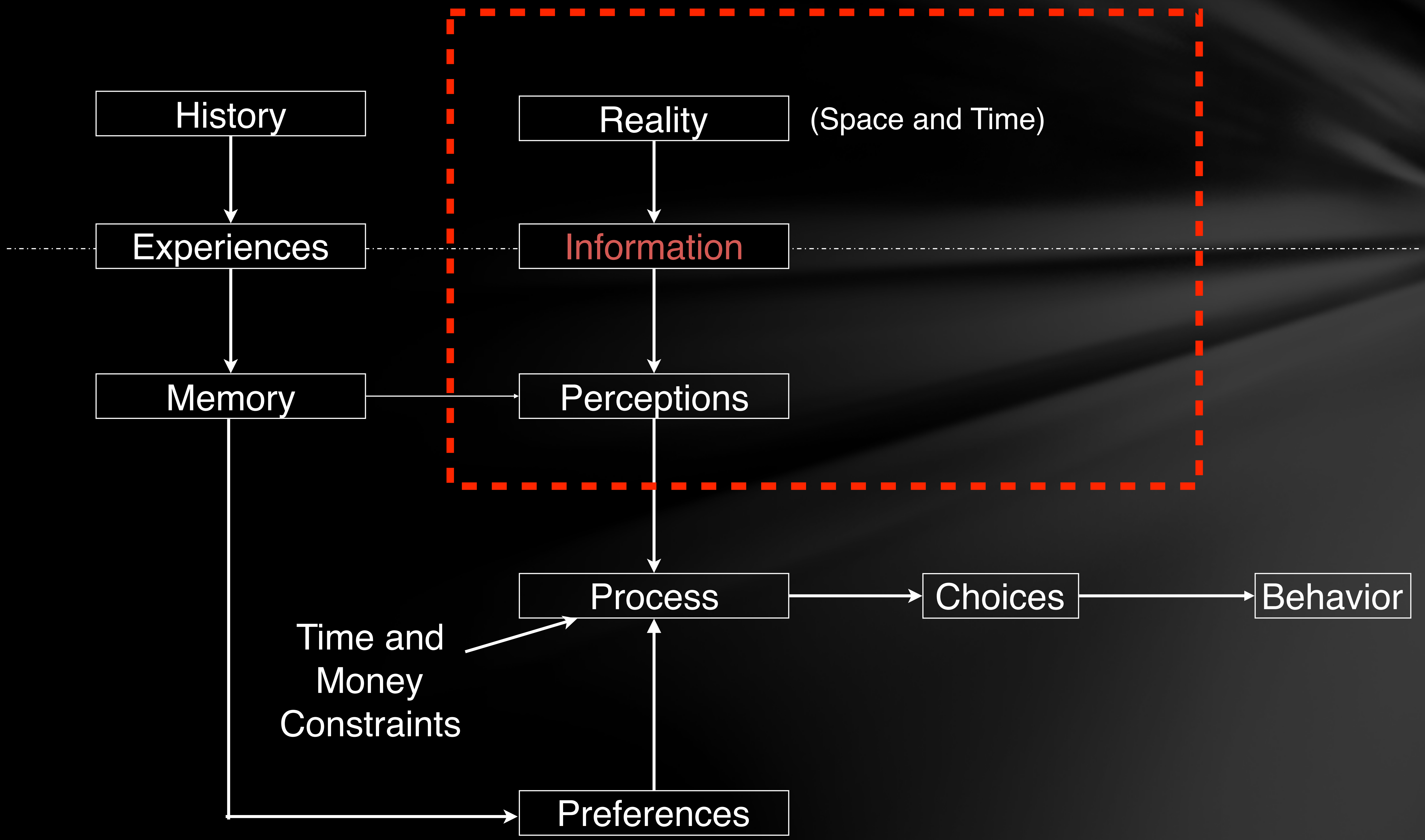


SECTION III. MOBILITY POLICY: FORMULATE, IMPLEMENT AND EVALUATE



Behavior — Theory — Policy

A Theory of Behavioral Choice



Theoretical Challenges in DNN

- Approximation
- Optimization
- Generalization

Poggio, T., Banburski, A., & Liao, Q. (2019). Theoretical Issues in Deep Networks: Approximation, Optimization and Generalization. arXiv preprint arXiv:1908.09375.

Correlation vs. Causality

Umbrella

CAUSAL HIERARCHY

Level	Typical Activity	Typical Question	Examples
 Association $P(y \mid x, c)$	Seeing random sampling	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
 Intervention $P(y \mid do(x), c)$	Doing random experimentation	What if? What if I do X?	What if I take aspirin, will my headache be cured?
 Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospecting	What if I had acted differently? Why?	Was it the aspirin that stopped my headache?

Bareinboim and Pearl 2016 Causal Inference and the Data-fusion Problem

Machine Learning for Economics

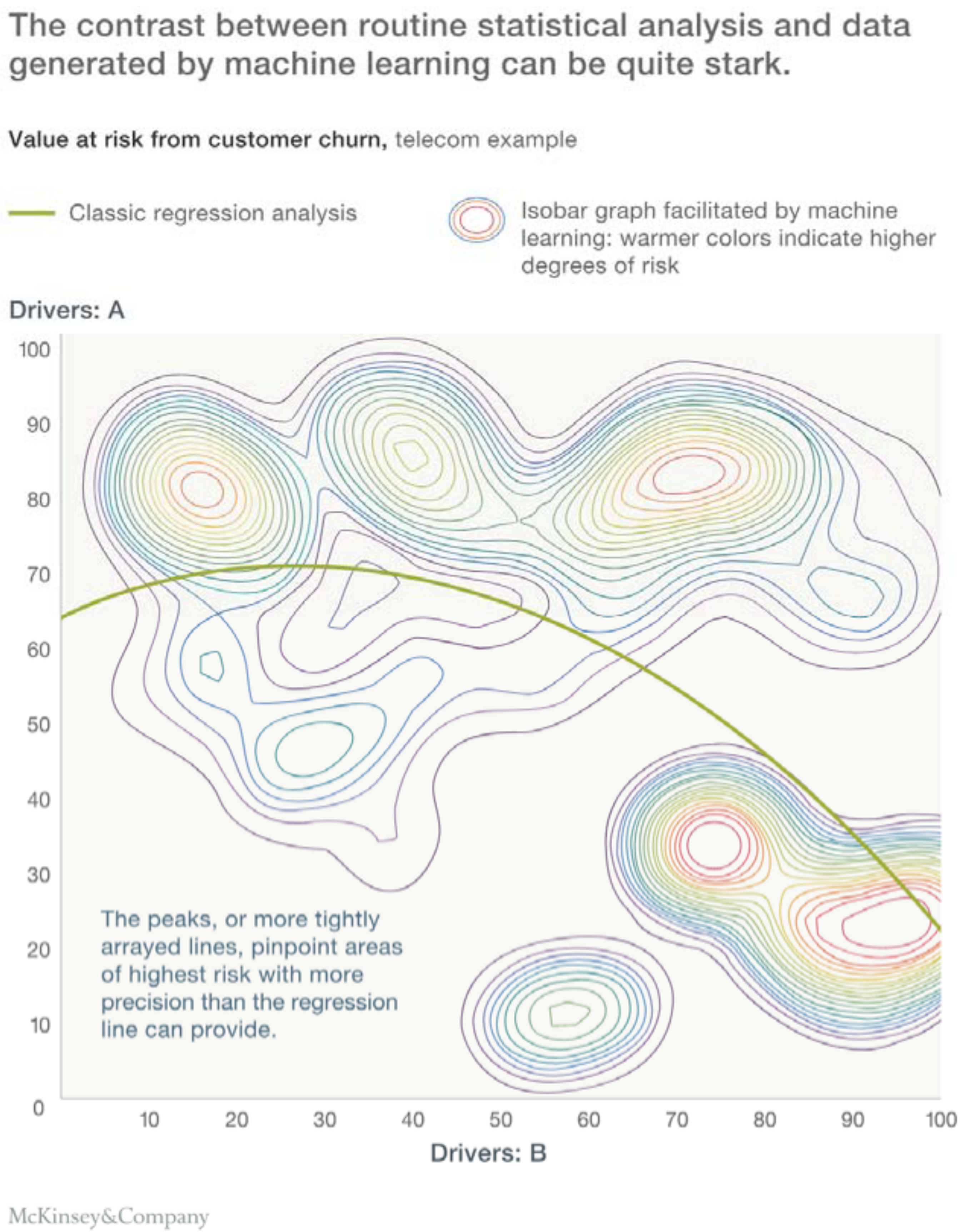
SUSAN ATHEY (STANFORD GSB)

What's New About ML?

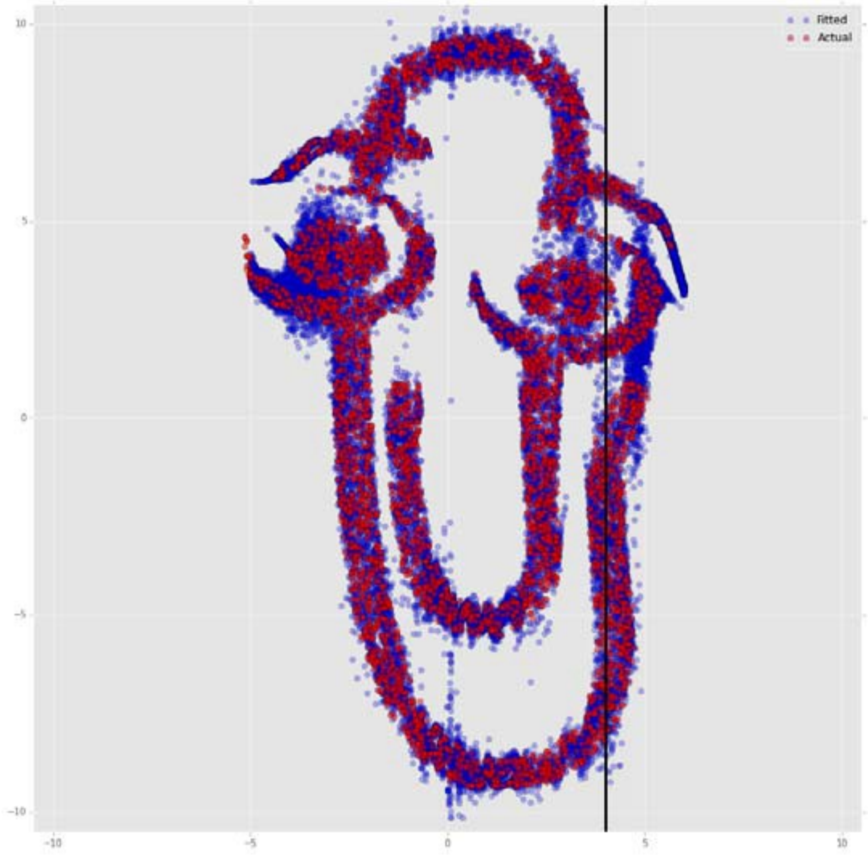
Flexible, rich, data-driven models

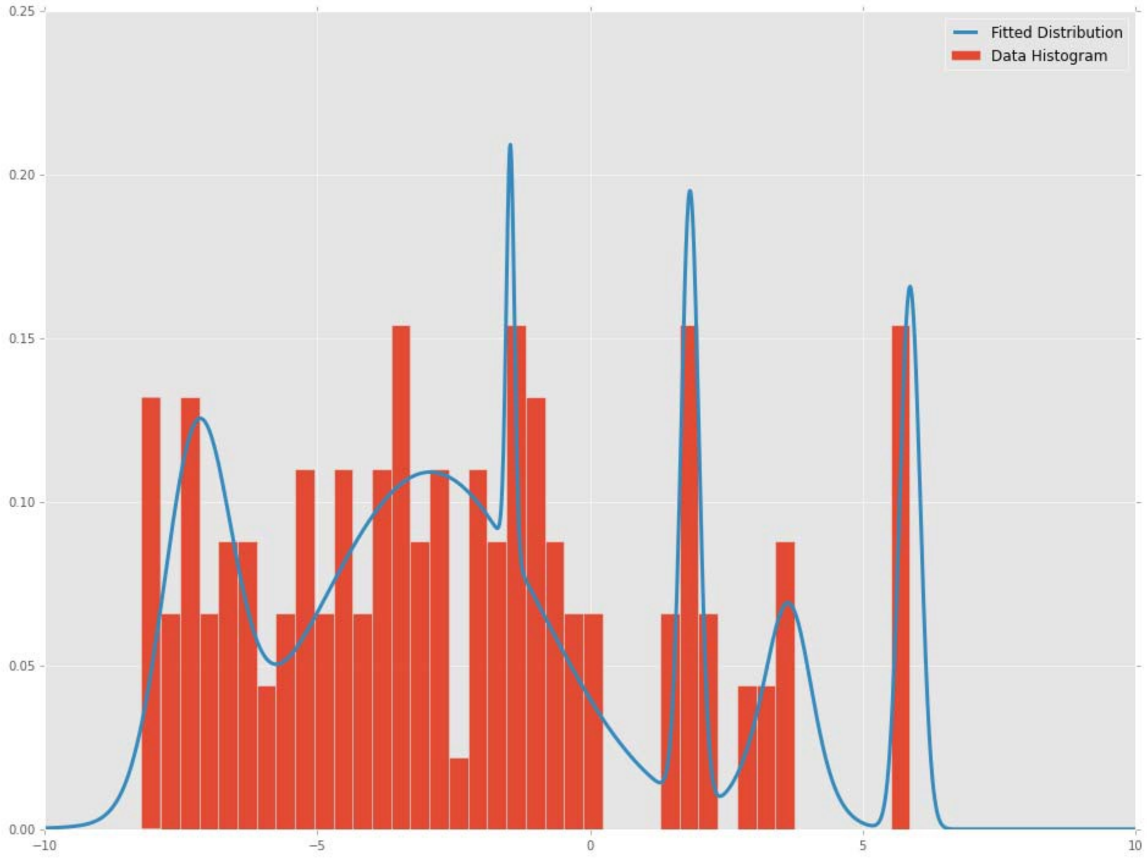
Increase in personalization and precision

Methods to avoid over-fitting



Ability to Fit Complex Shapes





Contrast with Traditional Econometrics

Economists have focused on the case with substantially more observations than covariates (N>>P)

- In-sample MSE is a good approximation to out-of-sample MSE
- OLS is BLUE, and if overfitting is not a problem, then no need to incur bias
- OLS uses all the data and minimizes in-sample MSE

OLS obviously fails due to overfitting when P~N and fails entirely when P>N

- ML methods generally work when P>N

Economists worry about estimating causal effects and identification

- Causal effects
- Counterfactual predictions
- Separating correlation from causality
- Standard errors
- Structural models incorporating behavioral assumptions

Identification problems can not be evaluated using a hold-out set

- If joint dist'n of observable same in training and test, will get the same results in both

Causal methods sacrifice goodness-of-fit to focus only on variation in data that identifies parameters of interest

What We Say v. What We Do (Econometrics)

What We Say

- Causal inference and counterfactuals
- God gave us the model
- We report estimated causal effects and appropriate standard errors
- Plus a few additional specifications for robustness

What we do

- Run OLS or IV regressions
  - Try a lot of functional forms
  - Report standard errors as if we ran only one model
  - Have research assistants run hundreds of regressions and pick a few “representative” ones
- Use complex structural models
  - Make a lot of assumptions without a great way to test them

Key Lessons for Econometrics

Many problems can be decomposed into predictive and causal parts

- Can use off-the-shelf ML for predictive parts

Data-driven model selection

- Tailored to econometric goals
  - Focus on parameters of interest
  - Define correct criterion for model
  - Use data-driven model selection where performance can be evaluated
- While retaining ability to do inference

ML-Inspired Approaches for Robustness

Validation

- ML always has a test set
- Econometrics can consider alternatives
  - Ruiz, Athey and Blei (2017) evaluate on days with unusual prices
  - Athey, Blei, Donnelly and Ruiz (2017) evaluate change in purchases before and after price changes
  - Tech firm applications have many A/B tests and algorithm changes

Other computational approaches for structural models

- Stochastic gradient descent
- Variational Inference (Bayesian models)

See Sendhil Mullainathan et al (JEP, AER) for key lessons about prediction in economics

- See also Athey (Science, 2017)

Empirical Economics in Five Years: Susan's Predictions

Regularization/data-driven model selection will be the standard for economic models

Prediction problems better appreciated

Measurement using ML techniques an important subfield

Textual analysis standard (already many examples)

Models will explicitly distinguish causal parts and predictive parts

Reduced emphasis on sampling variation

Model robustness emphasized on equal footing with standard errors

Models with lots of latent variables

Knowledge Discovery

So How Do Computers Discover New Knowledge?

1. **Symbolists**--Fill in gaps in existing knowledge

2. **Connectionists**--Emulate the brain

3. **Evolutionists**--Simulate evolution

4. **Bayesians**--Systematically reduce uncertainty

5. **Analogizers**--Notice similarities between old and new

SRC: Pedro Domingos ACM Webinar Nov 2015 http://learning.acm.org/multimedia.cfm

The Five Tribes of Machine Learning

Tribe	Origins	Key Algorithm
Symbolists	Logic, philosophy	Inverse deduction
Connectionists	Neuroscience	Backpropagation
Evolutionists	Evolutionary biology	Genetic programming
Bayesians	Statistics	Probabilistic inference
Analogizers	Psychology	Kernel machines

SRC: Pedro Domingos ACM Webinar Nov 2015 http://learning.acm.org/multimedia.cfm

Representation, Evaluation and Optimization

- Representation
  - Probabilistic logic (e.g., Markov logic networks)
  - Weighted formulas → Distribution over states
- Evaluation
  - Posterior probability
  - User-defined objective function
- Optimization
  - Formula discovery: Genetic programming
  - Weighted learning: Backpropagation

SRC: Pedro Domingos ACM Webinar Nov 2015 http://learning.acm.org/multimedia.cfm
http://www.ibtimes.co.uk/elon-musks-1bn-non-profit-launches-gym-train-ai-atari-games-1557362

MIT Subjects Related to ML

- 6.036 Introduction to Machine Learning
- 6.862 Applied Machine Learning (6.862 = 6.036 + Term Project)
- 6.867 Machine Learning
- 9.520/6.860: Statistical Learning Theory and Applications
- 9.S914: Mathematical Statistics: A Non-Asymptotic Approach
- 6.883: Online Methods in Machine Learning: Theory and Applications

MIT Subjects Related to ML

- 4.S42 Machine Learning for Creative Design
- 6.S897/HST.S53: Machine Learning for Healthcare
- MAS.533 AI for Impact ~ Towards Health & Sustainability from People to Planet
- 6.268 Network Science and Models
- 11.s938/11.s196 Deep Learning for Urban Transportation

Deep Learning for Urban Transportation

Part I: Introduction and DNN Basics
- Sep 10 Introduction: Deep Learning Meets Transportation

Part II: Passenger and Traffic Flow Prediction
- Sep 17 Demand prediction and Deep learning basics
- Sep 24 Advanced modeling techniques: ConvLSTM, Attention, individualized predictions
- Oct 1 Guest Lecture Prof. Justin Dauwels
- Oct 8 Advanced Applications: Generative models (VAE), graph embeddings,

Part III: DNN and Demand Analysis
- Oct 22 DNN and Discrete Choice 1
- Oct 29 DNN and Discrete Choice 2 and Guest Lecture by Prof. Hai Wang
- Nov 5 DNN and Prospect Theory

Part IV: Reinforcement Learning and Control in Transportation
- Nov 12 RL and Control Part 1
- Nov 19 RL and Control Part 2
- Nov 26 Guest Lecture Prof. Cathy Wu

Part V: FAT and Summary
- Dec 3 Fairness, Accountability and Transparency
- Dec 10 Student Presentation