Context-Aware Online Traffic Prediction

Jie Xu, Dingxiong Deng, Ugur Demiryurek,
Cyrus Shahabi, Mihaela van der Schaar

University of California, Los Angeles
University of Southern California

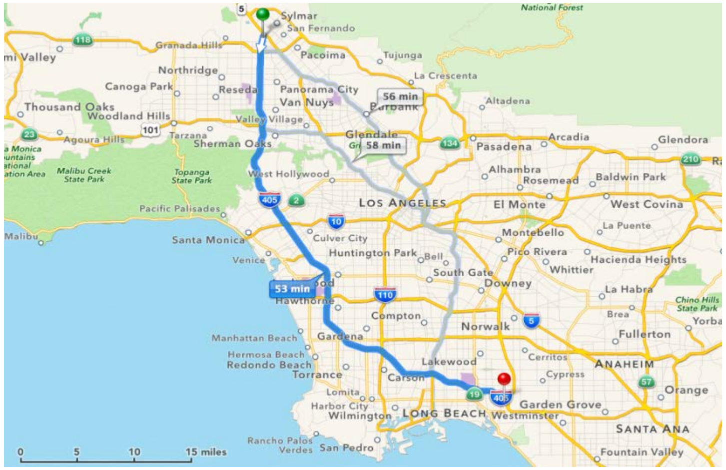
J. Xu, D. Deng, U. Demiryurek, C. Shahabi and M. van der Schaar, "Mining the Situation: Spatiotemporal Traffic Prediction with Big Data," *IEEE Journal of Selected Topics in Signal Processing (JSTSP) - Special Issue on Signal Processing for Big Data,* 2015.
J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. van der Schaar, "Context-Aware Online Spatiotemporal Traffic Prediction," accepted and to appear in *ICDM Workshop on Spatial and Spatio-Temporal Data Mining, 2014*

Online Traffic Prediction

- Traffic congestion - time and energy wasted
  - 4.2 billion vehicle-hours of delay
  - 2.8 billion gallons in wasted fuel
  - $87.2 billion in lost productivity (0.7% GDP)

- Use Big Data to enable intelligent transportation
  - Road sensor instrumentations
  - Auxiliary commodity sensors (CCTV cameras, GPS)

Traffic congestion on I-405



2000 sensors along 50 miles of I-405

Importance of Traffic Prediction

- Drivers: avoid congested areas
  - E.g. through intelligent navigation systems
- Policy makers: decide traffic regulations
  - E.g. replace a carpool lane with a toll lane
- Urban planners: design better pathways
  - E.g. adding new lanes, carpool lanes
- Civil engineers: plan construction zones
  - E.g. predict how short-term construction impacts traffic

Requirements and challenges

- **Google Maps**
  - Suggests current routes based on current traffic condition
  - Traffic is changing

Forecast future traffic based on current traffic

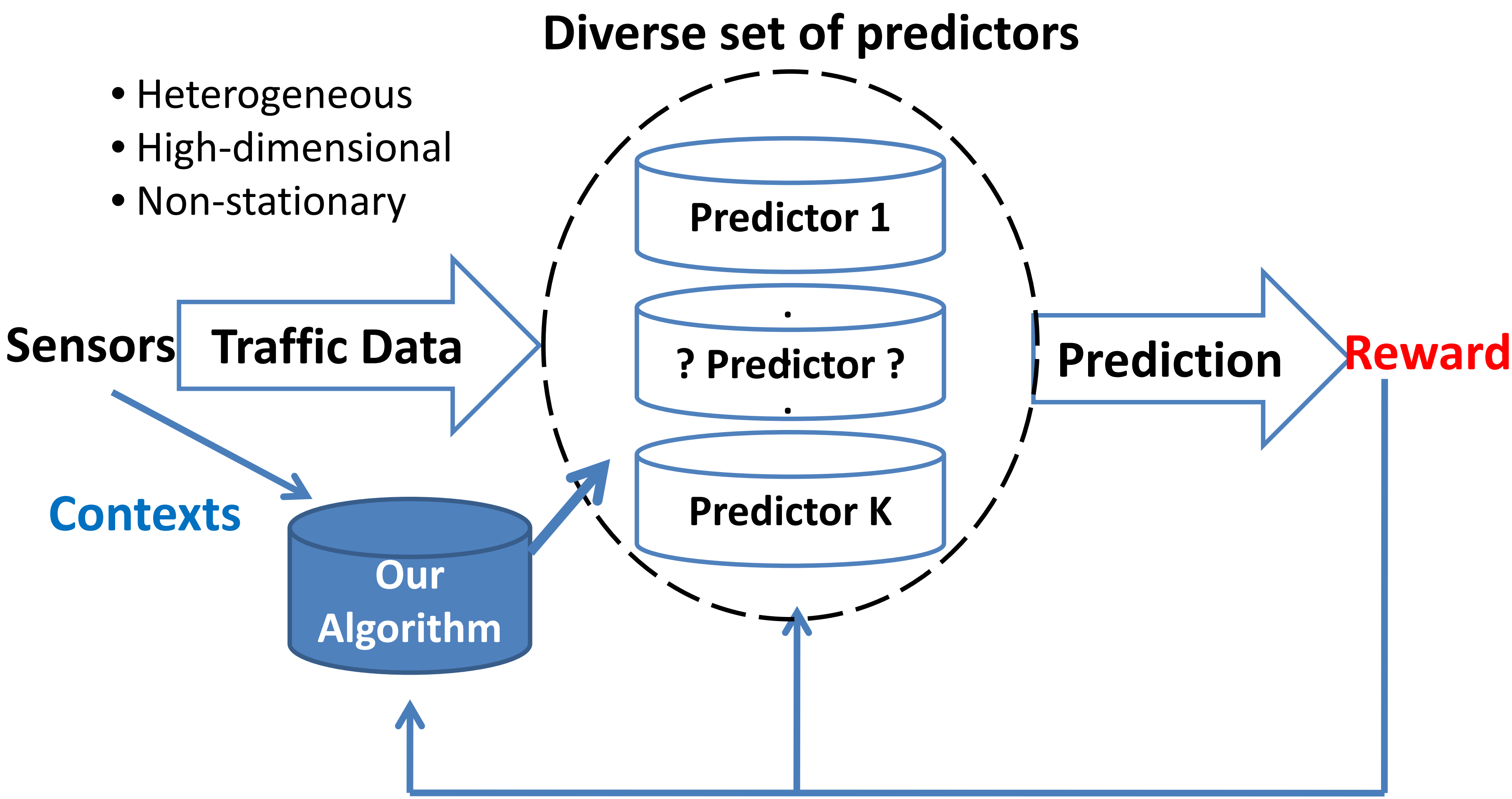
Predict accident-prone areas

- **Existing works on traffic prediction**
  - Time series anal, Naïve Bayes, Decision Tree, Nearest Neighbor etc.
  - Work well only in their envisioned situations
    - Typical conditions, rush hours, presence of accidents

- **Central challenges**
  - Many situations
  - Traffic situations are never identical
  - Traffic situations may be similar, but similarity must be discovered
  - High performance

Handle diverse traffic situations

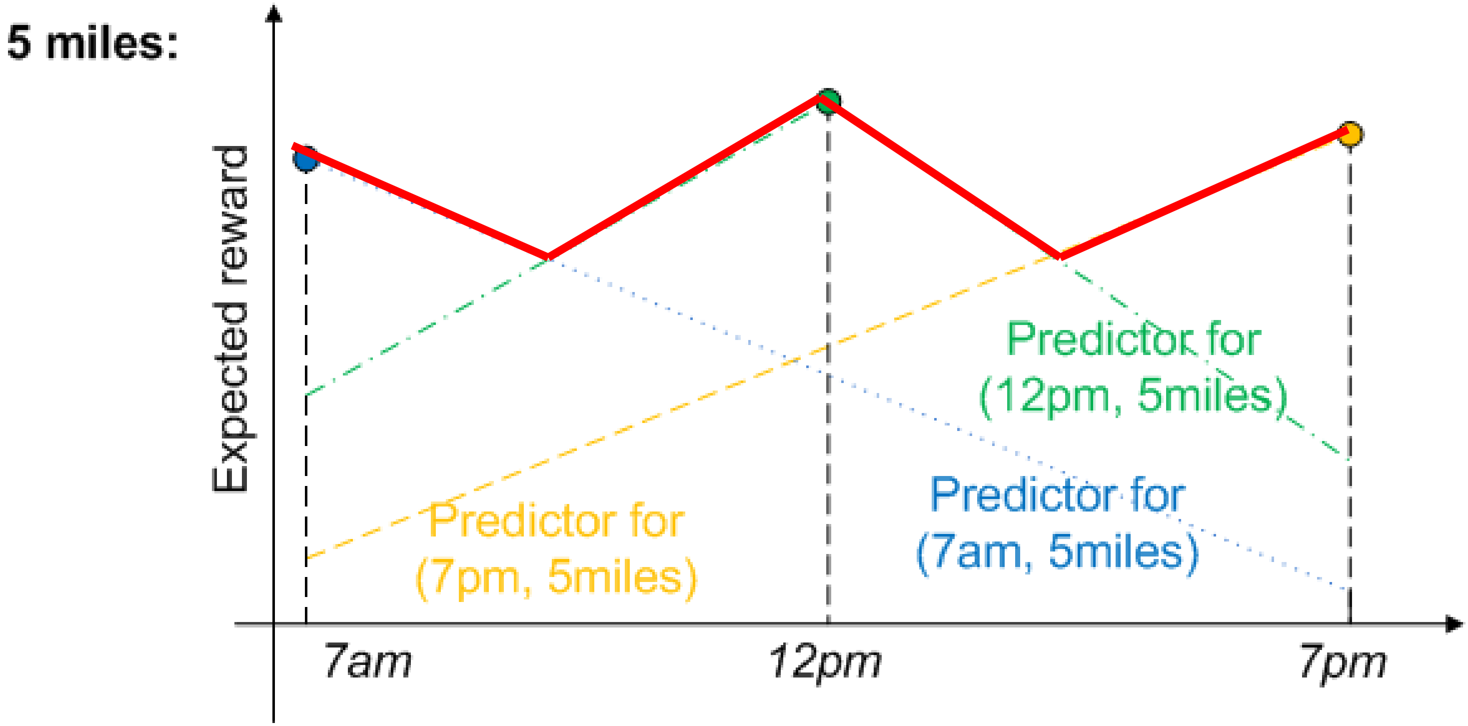
Our Approach for Traffic Prediction



Context – a D-dimensional vector: time of day, location, weather condition, traffic direction, road type, intersection, accident, etc.

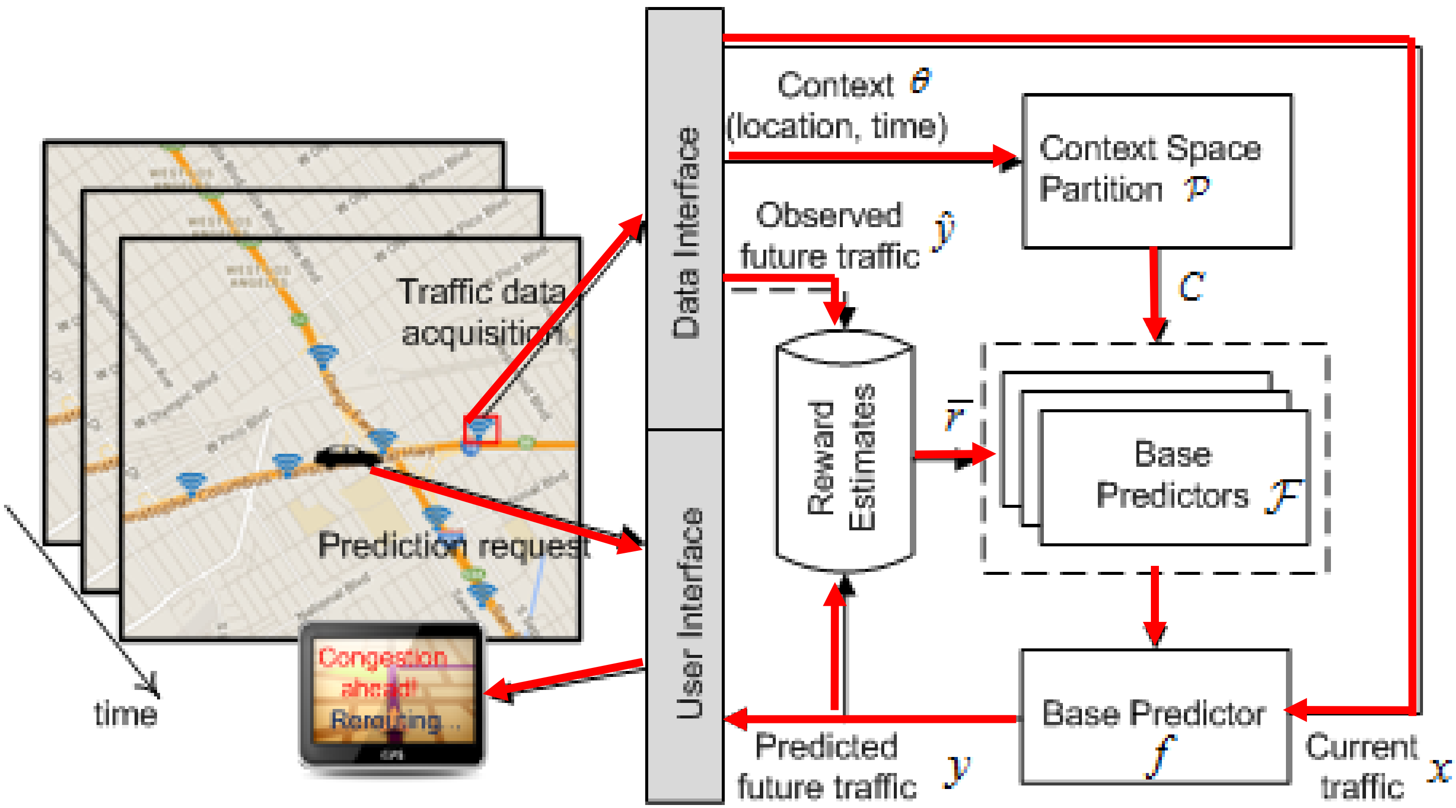
Our Approach for Traffic Prediction

– Predictor performance is <u>unknown</u> in different situations!
– Learn which predictor or combination of predictors to use.

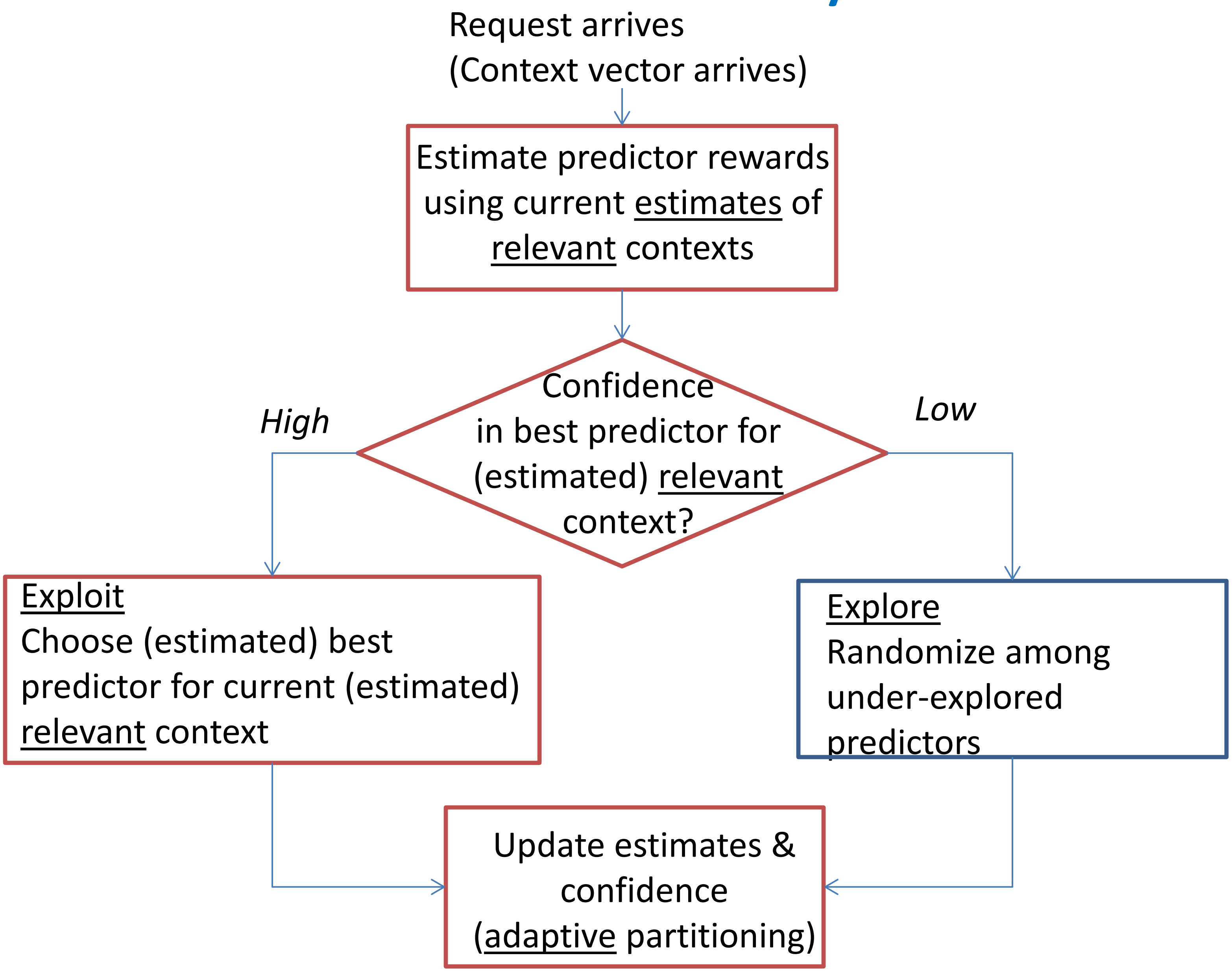


- Contexts organize information and are key to learning effectively ☺
- However, context space can be very large ☹
  - We discover what contexts are *<u>relevant</u>* for each prediction

Envisioned Traffic Prediction System



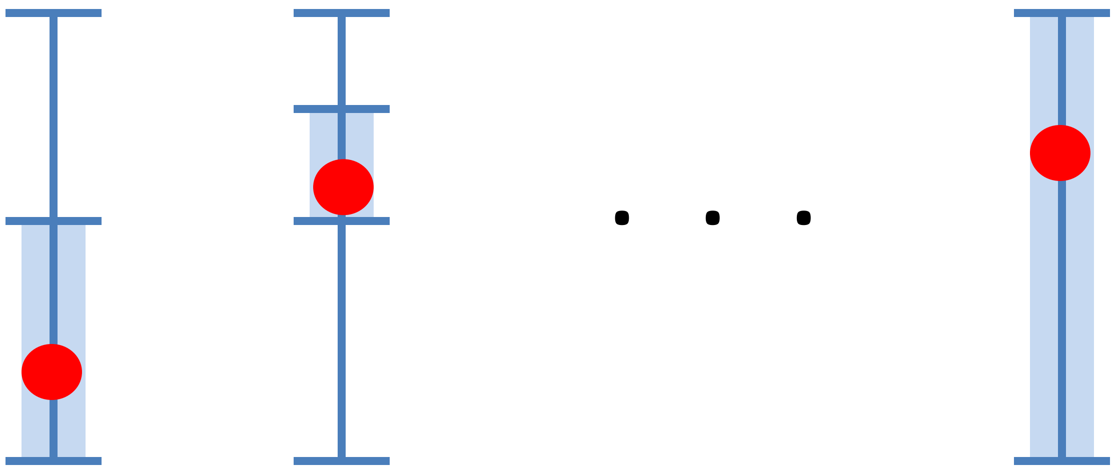
Context-Aware Discovery and Decision



Adaptive Partitioning

Current partitioning at time t

Arrival & context at time t



-> Choose a Predictor (explore/exploit)

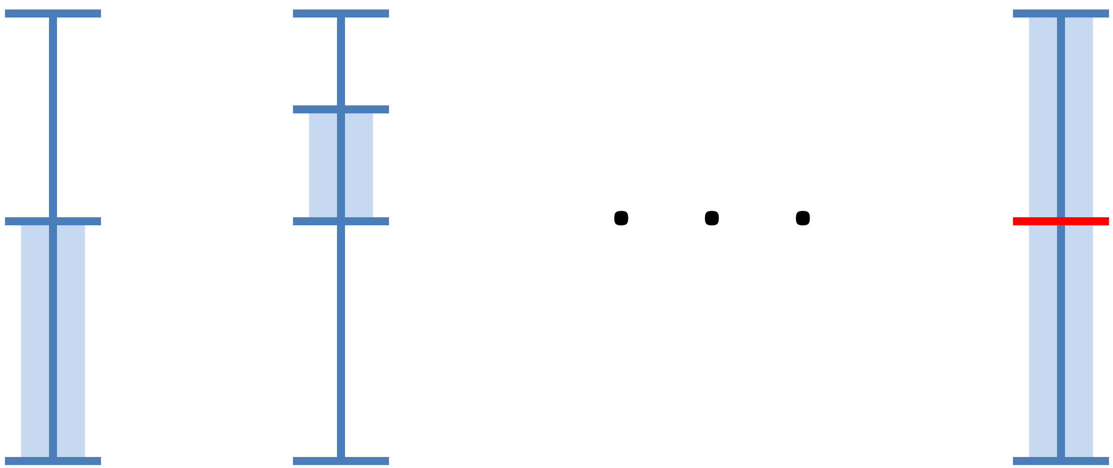
-> Observe Reward

Adaptive Partitioning – Updating

For each context is information above noise level?

No → No adaptation (partition remains the same)

Yes → Adaptation (subdivide one bin in partition)



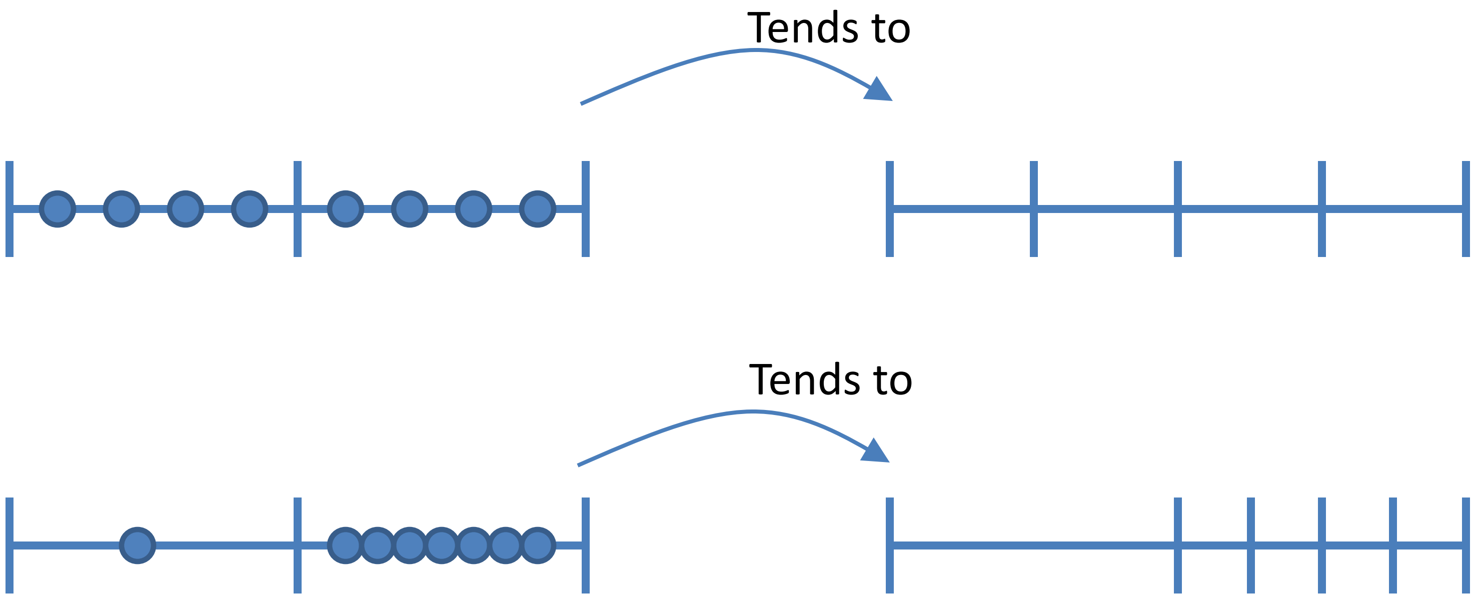
$$|\mu(a, x_{R(a)}) - \mu(a, x'_{R(a)})| \le L \cdot |x_{R(a)} - x'_{R(a)}|$$

Error estimate $\le L \cdot$ Length of active partition = Noise

Subtle Point

! Context arrival process matters:

for partitioning, estimates, decisions, performance



Learning performance

- **Regret** Reg(T): Cumulative (expected) loss up to time T from suboptimal action choices.
- Reg(T)/T: Rate of learning (to the optimal average reward).

- Regret is connected to *speed* of convergence.

Bounded	**Logarithmic**	**Sublinear**	**Linear**
$\mathrm{Reg}(T) < C$	$\mathrm{Reg}(T) < C \log T$	$\mathrm{Reg}(T) < CT^{\gamma}$ $\gamma < 1$	$\mathrm{Reg}(T) \sim CT$
Extremely fast convergence	Very fast convergence	Convergence	Non-convergence

Theoretical Results

Theorem: worst-case regret bound

$$\mathrm{Reg}(T) = \tilde{O}(T^{(2D_{rel}+3)/(2D_{rel}+4)})$$

– Optimal performance can be asymptotically achieved

$$\lim_{T \to \infty} \mathrm{Reg}(T)/T = 0$$

Better bounds can be derived depending on the arrival process

Experiments

- Online traffic data
  - Traffic sensor data from 9300 traffic loop-detectors
  - 5400 miles
  - Traffic parameters (occupancy, volume, speed) – one reading per sensor per minute
  - Traffic incident data (severity, location, type etc.)

- Evaluation Method
  - Different spatial settings
    - 100 locations, around 300 requests for each location.
  - Predict traffic congestion upon each prediction request
    - Congestion: traffic speed drops below a threshold

Performance

- Overall accuracy (with different congestion threshold)

	CA-Traffic	MU	AU	GDU
$\lambda = 50$ mph	0.94	0.83	0.83	0.82
$\lambda = 30$ mph	0.91	0.82	0.80	0.78

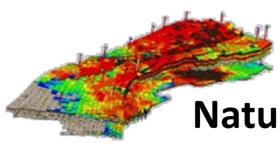
- Break-out accuracy

$(\lambda = 50\text{mph})$	**CA-Traffic**	MU	AU	GDU
10am	0.93	0.81	0.85	0.83
2pm	0.93	0.86	0.81	0.80
5pm	0.99	0.86	0.87	0.88
$(\lambda = 30\text{mph})$	**CA-Traffic**	MU	AU	GDU
10am	0.93	0.83	0.83	0.81
2pm	0.91	0.80	0.83	0.82
5pm	0.99	0.81	0.85	0.83

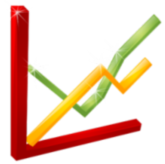
Benchmarks: Weighted Majority Methods

Additive Update (AU), Multiplicative Update (MU), Gradient Descent Update (GDU)

Methods - useful in **Many Applications**



Natural Systems
- Seismic monitoring
- Wildfire management
- Water management



Stock Markets
- Impact of weather on securities prices
- Analysis of market data at ultra-low latencies

Virology

Personalized e-Learning

Smart Cities

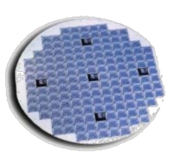
Energy

Transportation
- Intelligent traffic management
- Collision forecasting

Cyber-security

- Real time attack detection
- Real time attack prevention

Social media
- Popularity forecasting



Manufacturing
- Process control for microchip fabrication

Healthcare Informatics
- Remote healthcare monitoring
- Long-term care
- Assist-diagnosis tools
- Many patients: screening, clinical trials

Telecom
- Infrastructure adaptation
- Real-time services, billing, advertising