Pareto Optimal Streaming Unsupervised Learning

Soumya Basu, Steven Gutstein*, Brent Lance* and Sanjay Shakkottai

The University of Texas at Austin, *Army Research Lab

Streaming Unsupervised Learning

Tasks: Images arriving online for classification

Agents: Human agents, Neural-net classifiers

- **Different expertise**: Unknown confusion matrices
- **Deterministic labeling**: Label for each image fixed
- **Limited Processing Power**: One image/round

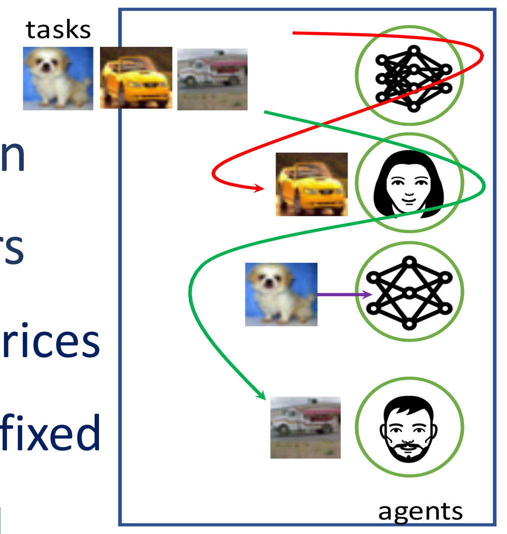
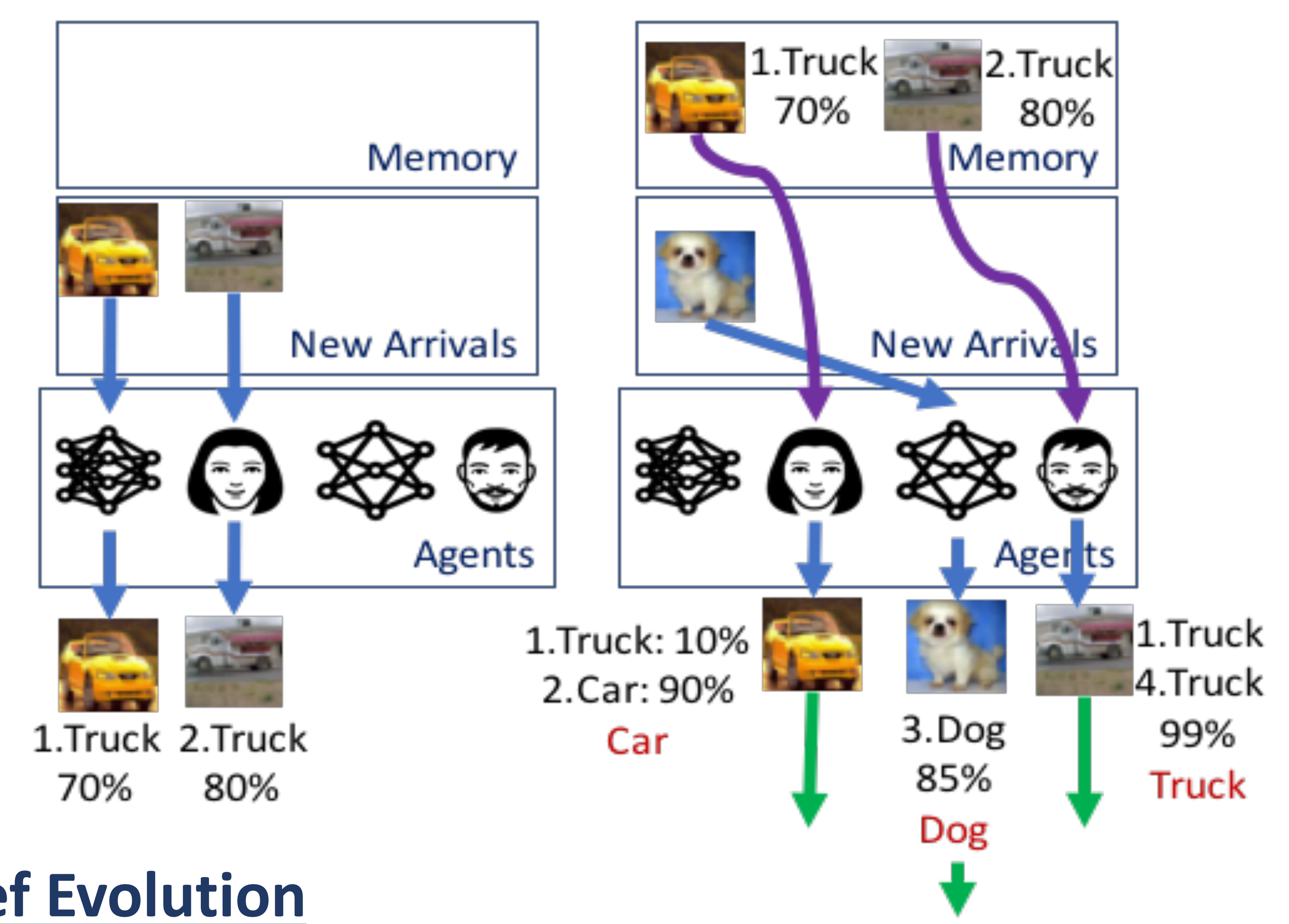


Image credits: CIFAR-10, A. Krizhevsky, 2009; thenounproject.com, (NNs - K. M. Synstad; Faces - A. Selimov)

Resource Allocation and Label Aggregation

- Sequential routing to subset of agents
- Collected labels are continually aggregated
- Aggregation using estimated confusion matrices



Belief Evolution

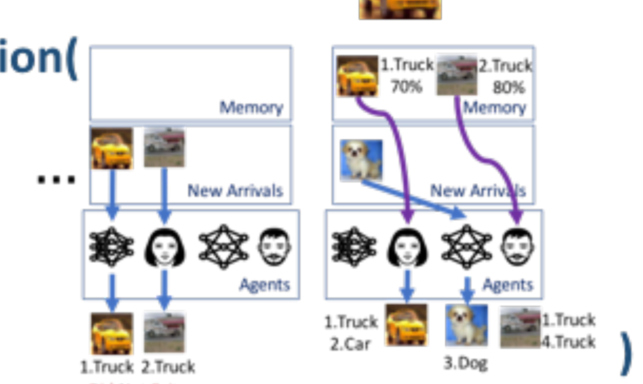
Causal Policy: Scheduling, aggregation, and labeling functions of history

How does the belief of each image evolve under a causal policy?

- Potential dependencies across images due to correlated decisions

Probability(True Label = Car |ENTIRE HISTORY)

= function(



)

We prove belief evolution is product form

∝ Probability(1.Truck | True Label = Car) ×

Probability(2.Car | True Label = Car)

Pareto Optimality

Accuracy vs Arrival Rate Tradeoff:

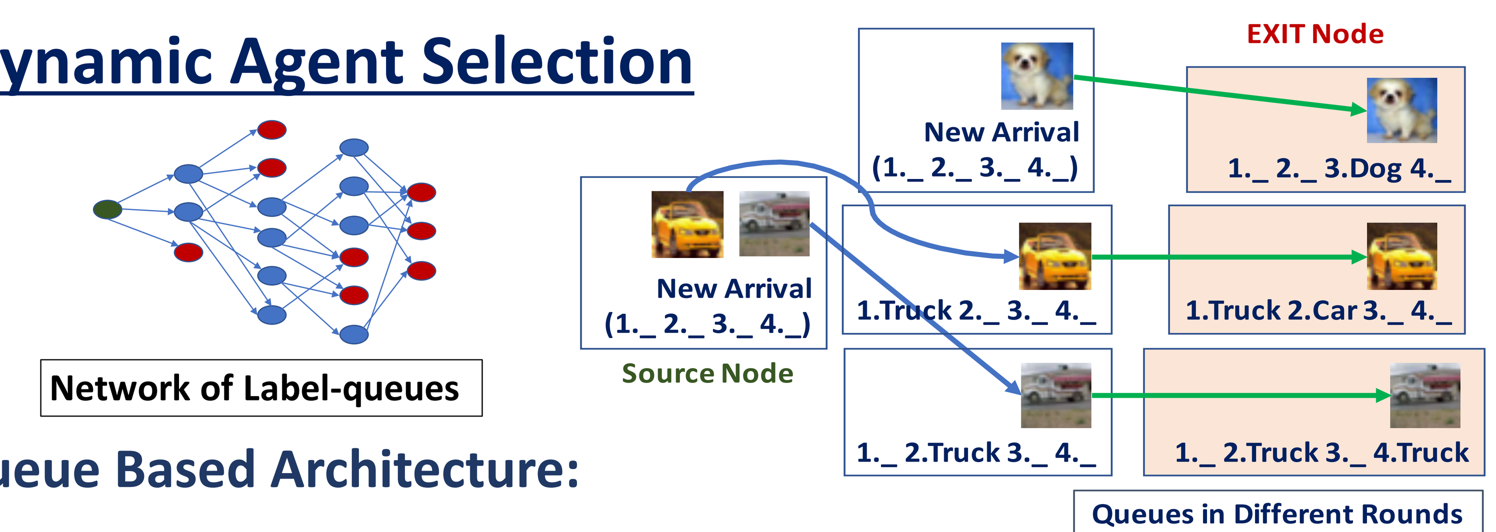
- Higher number of Labels leads to higher accuracy (product form)
- Higher number of labels per image implies smaller max arrival rate

Threshold Accuracy: For each exiting sample

Either **'All labels are collected'** Or **'Confidence > Threshold'**

Pareto Optimality: Operate anywhere in the Pareto curve with bounded memory in expectation

Dynamic Agent Selection



Queue Based Architecture:

- Network of interconnected label queues
- An image traverses queues as labels are collected

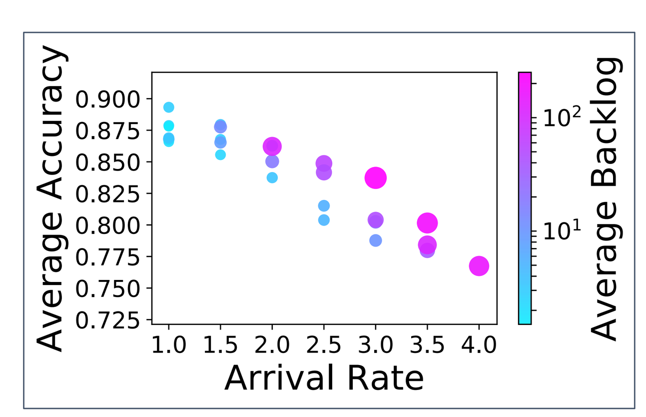
Bayesian Threshold Departure:

- **Exit nodes:** 'Complete labeling' or 'Confidence > Threshold'

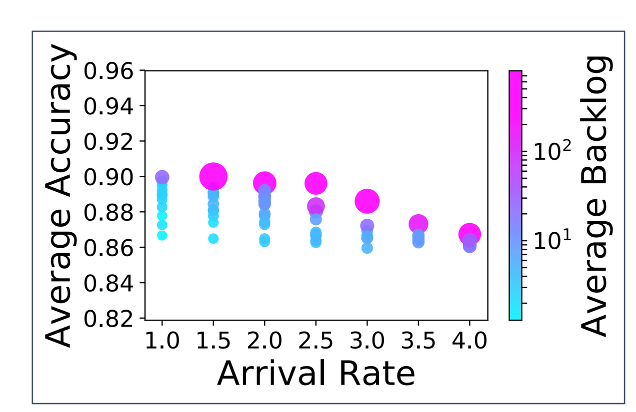
Max-weight Scheduling:

- Image-Agent matching: Max **expected back-pressure**

Experiments with Neural Network Ensembles



Pareto Region Exp 1



Pareto Region Exp 2

Experiment 1

- **6 Classifiers:** Three AlexNet, one VGG-19, and two ResNet-18
- **3 Labels:** Group 1 (airplanes, ships, trucks, cars), Group 2 (birds, frogs cats), and Group 3 (dogs, deer, horses)

Experiment 2

- **6 Classifiers:** One VGG-11 , one VGG-16, two VGG-19, and two ResNet-18
- **5 Labels:** Group 1 (airplanes, ships), Group 2(trucks, cars), Group 3(birds, frogs), Group 4 (cats, dogs) and Group 5(deer, horses)

Online Dawid-Skene Model

K classes and M classifiers

Arrival: I.i.d. $N(\tau)$ images arrive in round τ, **rate** $\mathbb{E}[N(\tau)]$**img/round**

Agent 'i': Confusion matrix $C_i(\cdot,\cdot)$

True label of Image 'j': T_j chosen with p.m.f. $p_g \in \Delta_K$

Label of Image 'j' from Agent 'i': $L_j(i)$ chosen from p.m.f. $C_i(T_j, \cdot)$

Image 'j': Deterministic (M+1)-tuple $(L_j(1), \ldots, L_j(M), T_j)$

System Dynamics

Length for label-queue ℓ in time τ, $Q_\ell(\tau)$

Image-Agent schedule in time τ, $S(\tau)$

New Labels provided by classifiers $L(\tau)$: determined by $S(\tau)$, agents

Departure from label-queue ℓ in time τ $D_\ell(\tau)$: determined by $S(\tau)$

Arrival in label-queue ℓ in time τ $A_\ell(\tau)$: determined by $S(\tau)$, $L(\tau)$

Expected Back-pressure:

- Expectation taken over new label which is random
- $S(\tau) = argmax \sum_\ell Q_\ell(\tau)(D_\ell(S) - \mathbb{E}[A_\ell(S, L)])$
- Time complexity # Agents × Sum of queue lengths

Online Learning

Explore: In round τ w.p. $\frac{\log(\tau)}{\tau}$, one unlabeled image to **ALL** agents

Tensor decomposition based approach to unsupervised learning [1]

- Offline tensor decomposition unsuited for running time

Online tensor decomposition using tensor power method (TPM) [2]

- Fixed number of initializations in TPM does not work
- Repeated Initializations until termination condition is met
- Constant amortized runtime per round

Performance Guarantees

For any **arrival rate** λ and **threshold** θ**, for any** $\epsilon > 0$,
if $(\lambda + \epsilon, \theta + \epsilon)$ lies in the Pareto region
then Max-weight scheduling + online tensor decomposition
support (λ, θ) with bounded expected memory.

References

[1] Zhang et al. 'Spectral methods meet EM: A provably optimal algorithm for crowdsourcing.'

[2] Anandkumar et al. 'Tensor decompositions for learning la- tent variable models'