Selective Experience Replay for Lifelong Learning

David Isele and Akansel Cosgun

Presenter: Ashique KhudaBukhsh

School of Computer Science

Carnegie Mellon University (CMU)

Pittsburgh, PA, USA

Methods and Challenge

- Challenge: catastrophic forgetting in DRL
- Method:
  - augment FIFO buffer with an episodic memory
  - Store selective experiences in the episodic memory for replay
- Research question: which experiences to store?
- Contribution:
  - Proposes four ranking functions for experiences
  - Two of which can deal with catastrophic forgetting

Background

- experience e = (s, a, s', r)
- Goal: maximize future discounted reward

$$R_t = \sum_{k=t}^{T} \gamma^{k-t} r_k$$

- Loss for an individual experience

$$\mathcal{L}(e_i, \theta) = \left(r_i + \gamma \max_{a'_i} Q(s'_i, a'_i; \theta) - Q(s_i, a_i; \theta) \right)^2$$

- Lifelong setting

$$\frac{1}{m} \sum_{j=1}^{m} \frac{1}{n_j} \sum_{i=1}^{n_j} \mathcal{L}(e_i, \theta)$$

Strategy: surprise

- *Found to be useful in neuroscience experiments with rodents*
- *Rank experiences by surprise*

$$\mathcal{R}(e_i) = |r_i + \gamma \max_{a'} Q(s'_i, a') - Q(s_i, a_i)|$$

Strategy: reward

- *Rank experiences by reward*

$$\mathcal{R}(e_i) = |R_i(e_i)|$$

Strategy: distribution matching

- *Goal: distribution of experience captures combined distribution of all tasks*
- Experiences are arriving sequentially
  - Down-sample in a way all tasks are equally likely to be stored in the episodic memory
- Maps into a reservoir sampling problem
- *K* reservoir slots *N* sequentially arriving experience
  - (*i* <= *K)* automatically gets stored
  - *(i > K)*, randomly sample *j* from [1, *i*]
    - If j < K, replace the j^{th} slot with the new experience
  - Every experience has K/N probability to be the reservoir slots after all experiences are observed

Strategy: coverage maximization

- *Rank experiences by neighbor-count*

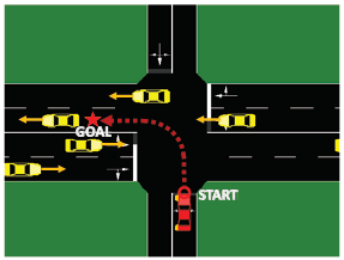
$$\mathcal{N}_i = \{e_j \ s.t. \ \text{dist}(e_i - e_j) < d\}$$
$$\mathcal{R}(e_i) = -|\mathcal{N}_i|, \text{order according to rank}$$

Data set

- Five intersection tasks

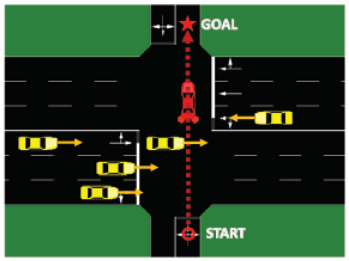
(a) *Right*

(b) *Left*



(c) *Left2*

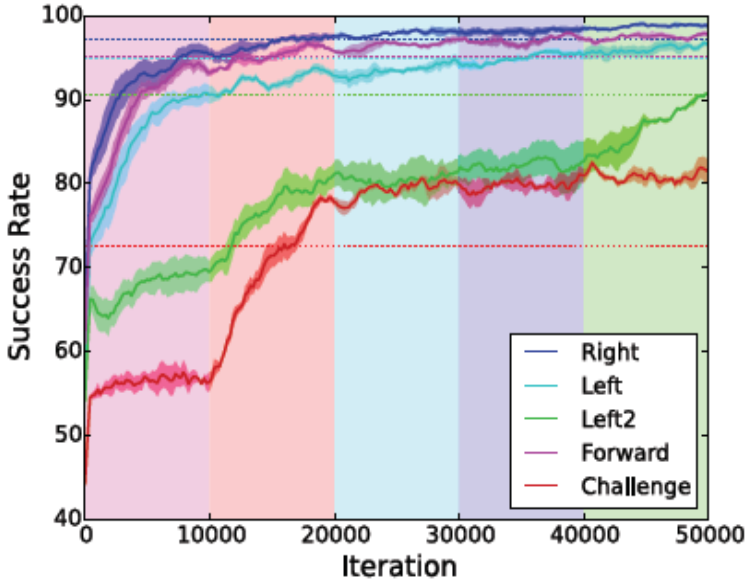
(d) *Forward*



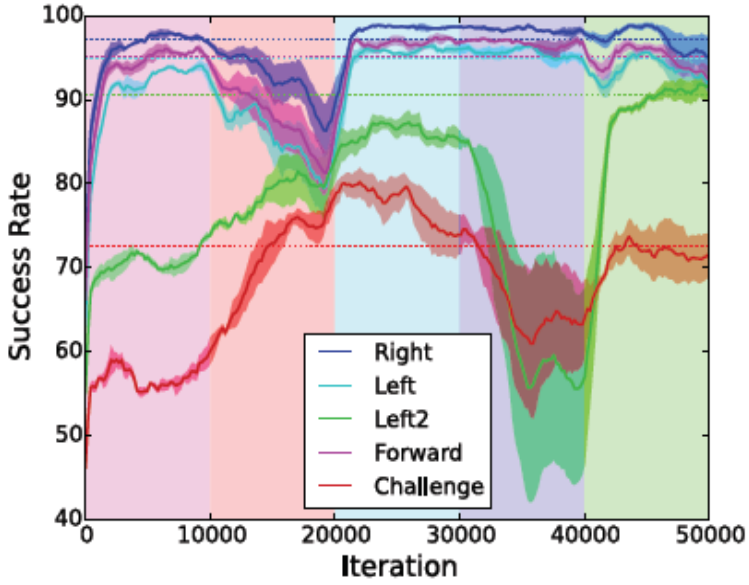
(e) *Challenge*

- One fixed order that exhibited catastrophic forgetting

Baseline



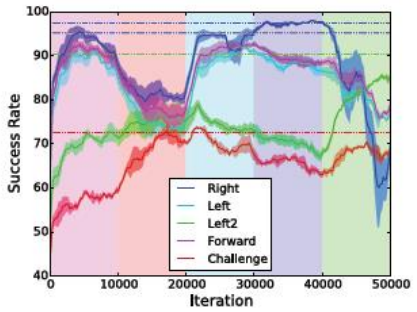
(a) Unlimited capacity



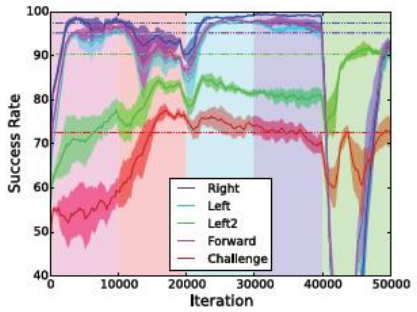
(b) Limited capacity (FIFO)

- Order: forward -> challenge -> left -> right -> left2

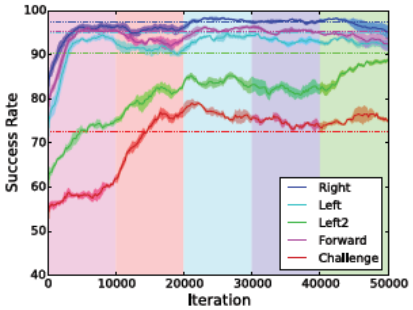
Relative Performance of Selection Strategies



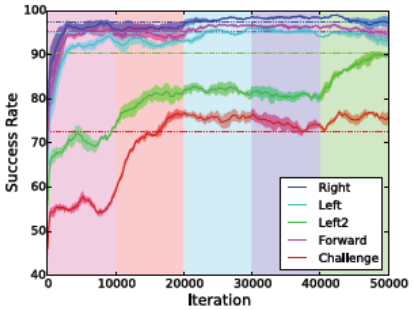
(a) Surprise



(b) Reward



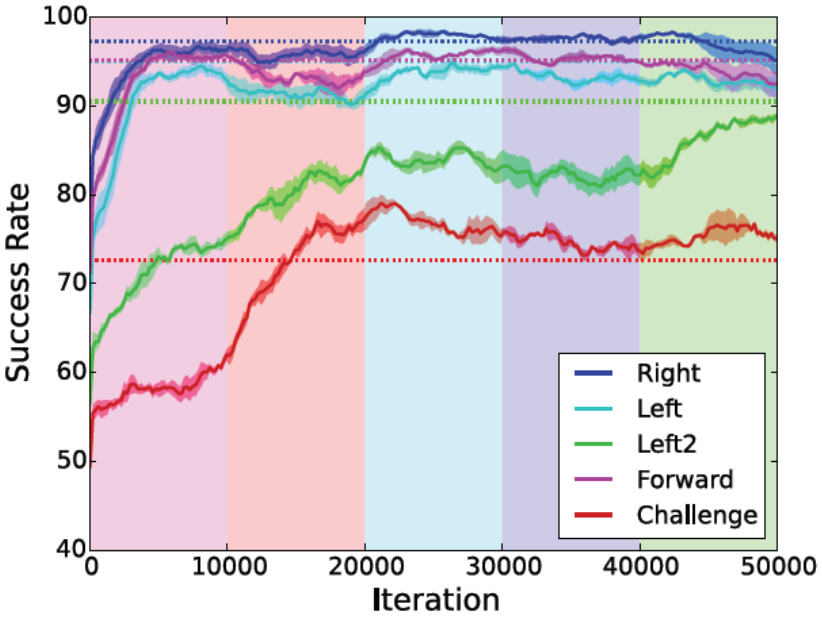
(c) Coverage maximization



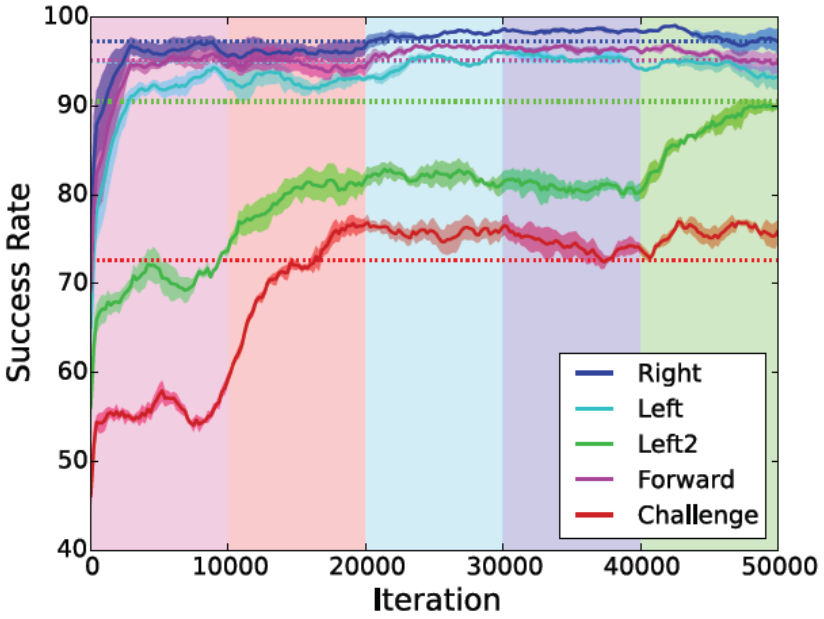
(d) Distribution matching

- Order: forward -> challenge -> left -> right -> left2
- Coverage maximization, distribution matching address catastrophic forgetting

Closer Look on the Strategies that Work

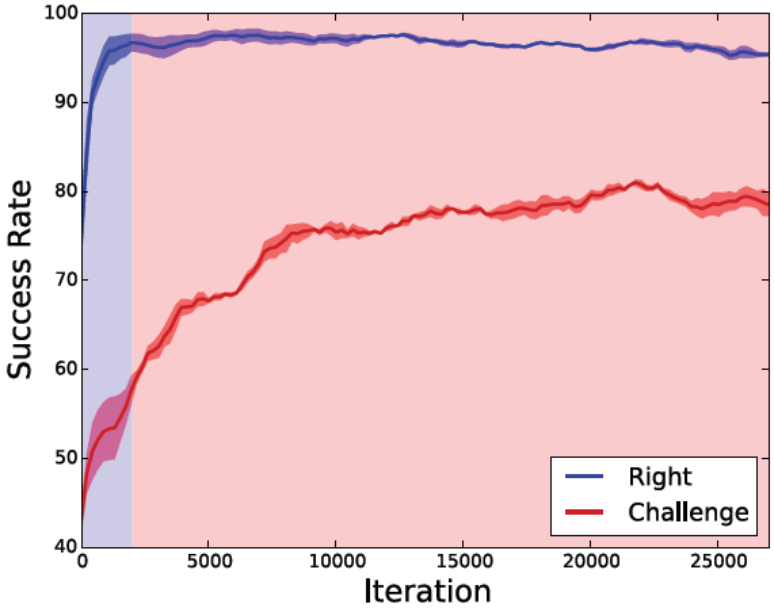


(c) Coverage maximization

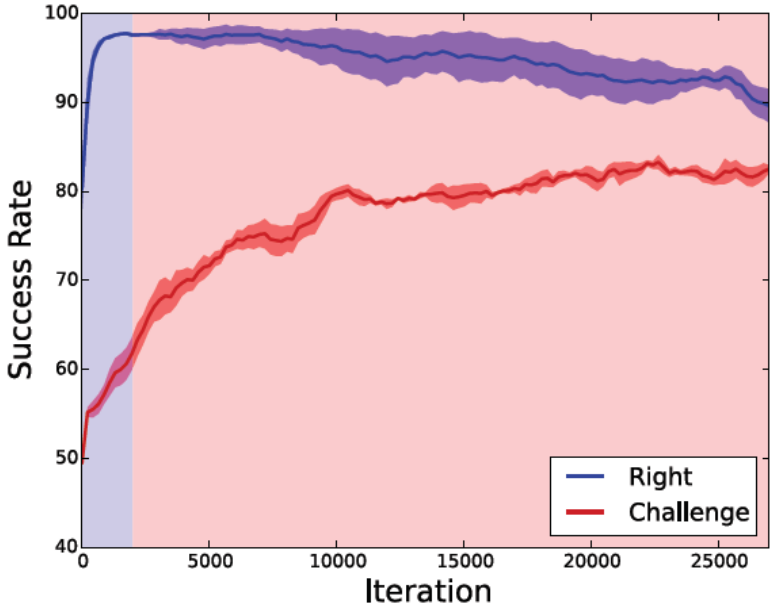


(d) Distribution matching

A Special Case where Coverage Works Better

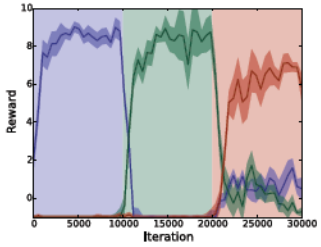


(a) Coverage Maximization

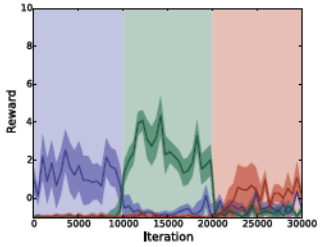


(b) Distribution Matching

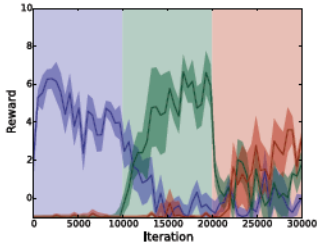
Grid World



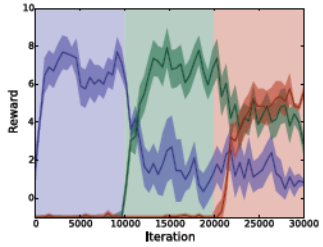
(a) No Selection



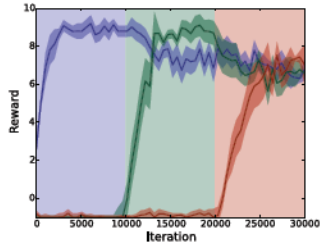
(b) Suprise



(c) Reward

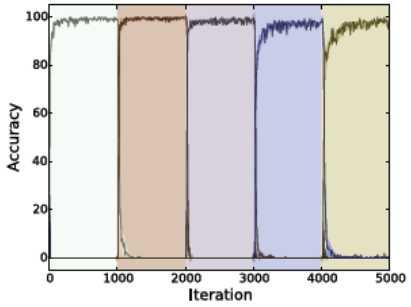


(d) Coverage Max.

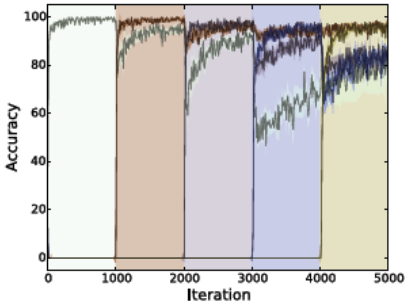


(e) Distribution Matching

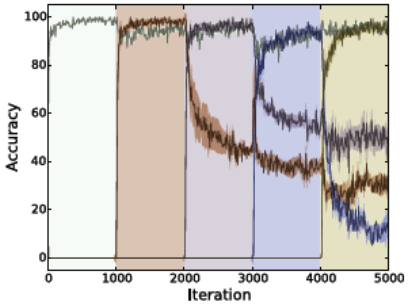
MNIST



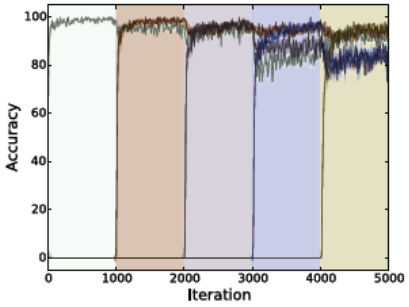
(a) No Selection



(b) Suprise



(c) Coverage Maximization



(d) Distribution Matching

Takeaways

- Episodic memory augmenting the FIFO buffer helps
- Distribution matching and coverage address catastrophic forgetting well
- Coverage works better when training is imbalanced