

Unsupervised CL

especially in Streaming Environments

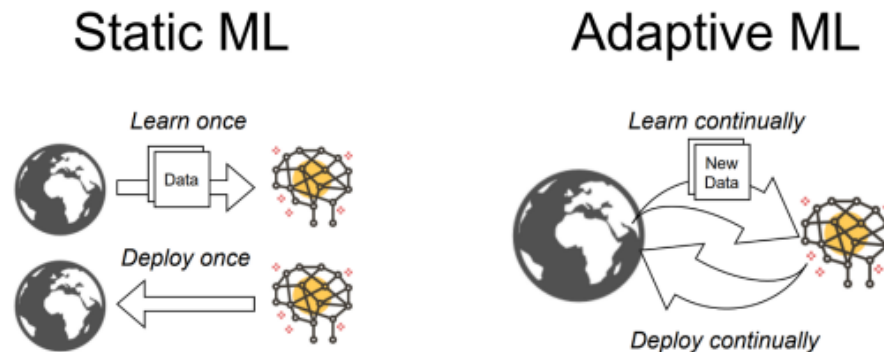
MinSeon Kim / 2022.09.23

Reading Points

- 어떠한 문제를 풀고자 하는지
- 어떤 기법을 채택 하였는지
- 실험의 구성은 어떻게 하였는지

Background Knowledge

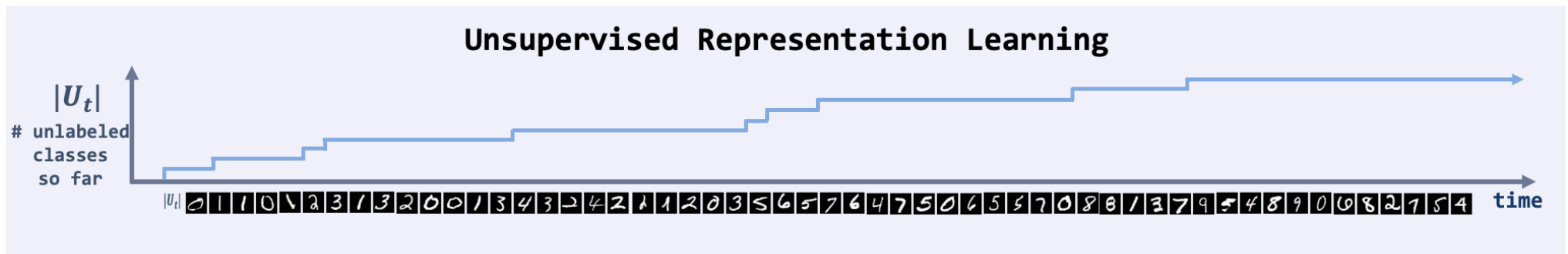
- Continual Learning 이란? : 시간에 따라서 학습 데이터가 주어지는 상황에서 Neural Network가 계속적으로 학습을 수행해 야되는 상황
- Problems to be addressed
1) Catastrophic Forgetting, 2) Limited Computing Resource (memory)



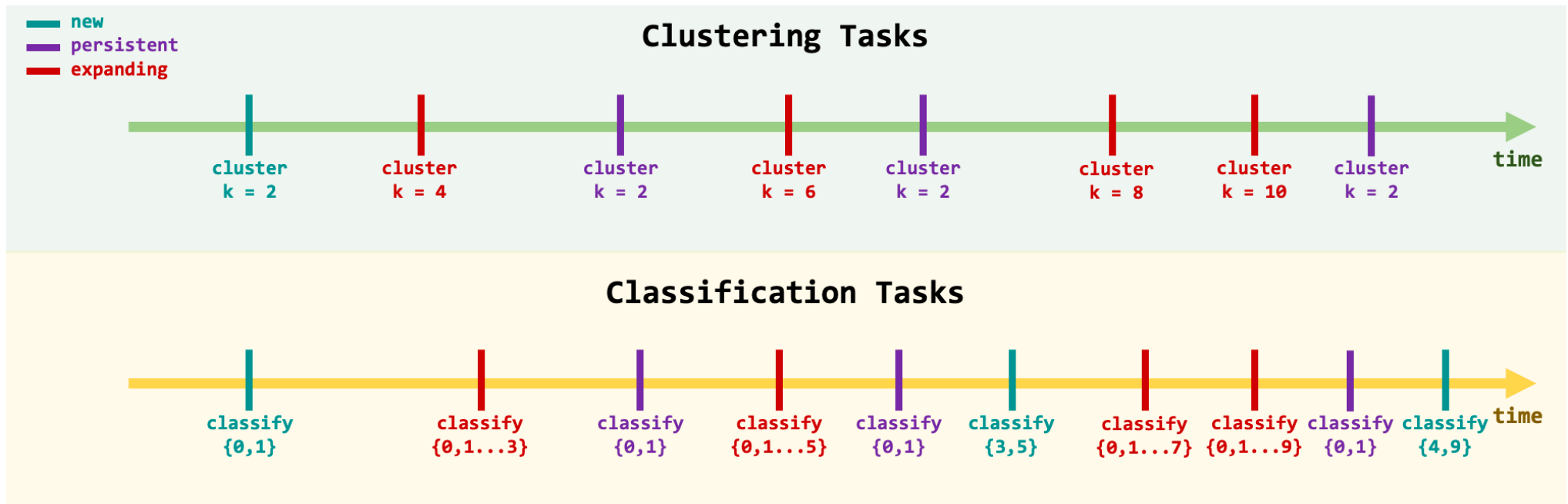
1. Unsupervised Progressive Learning and the STAM Architecture

ICML 2020

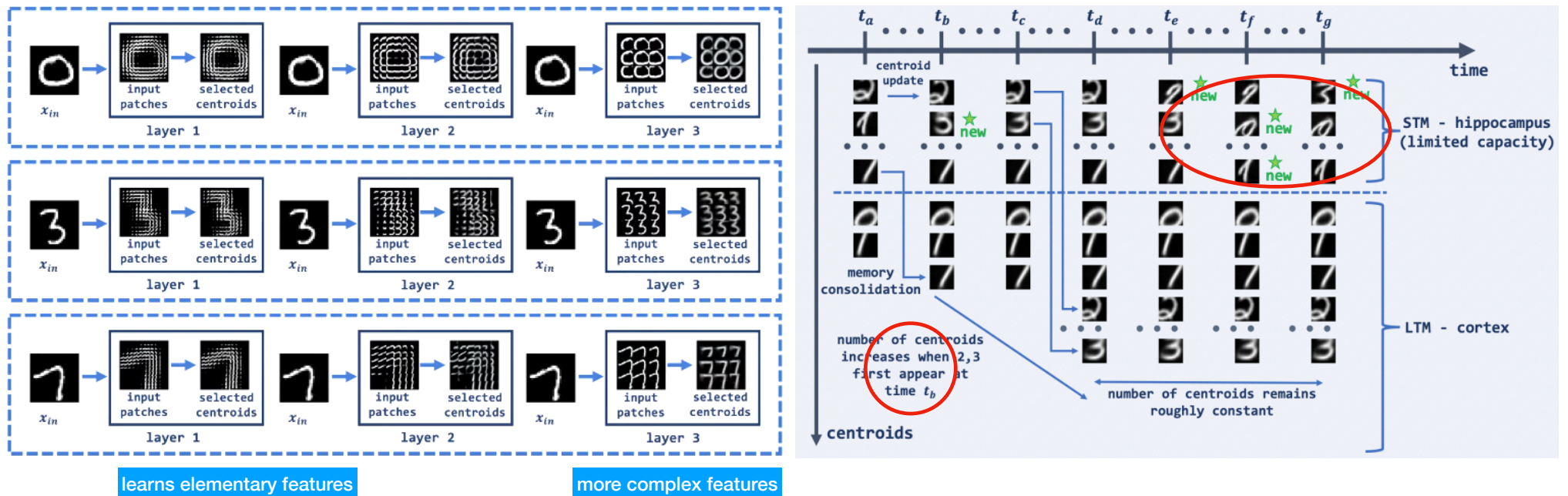
- UPL
 - No replay of previously seen data (제한된 메모리와 시간)
 - Non-stationary input: requires novelty detection & anomaly detection (event shift를 식별하기 위하여)
 - Memory is required (compact한 representation을 저장하여 미래의 task와 연관짓기 위해서)



- downstream task를 위한 useful representations를 학습하는것이 목표
- periodically하게 unsupervised clustering을 수행한다
- 제한된 개수의 라벨 데이터를 이용해 supervised classifications도 수행



- STAM("Self-taught Associative Memory")
: receptive field 사이즈를 증가해 나가면서 online clustering(centroid learning)을 수행하여 feature 학습
: dual-memory hierarchy (LTM & STM)을 통한 기존의 학습된 feature 보존
- Inputs are broken up into overlapping receptive fields and map to a bank of centroids



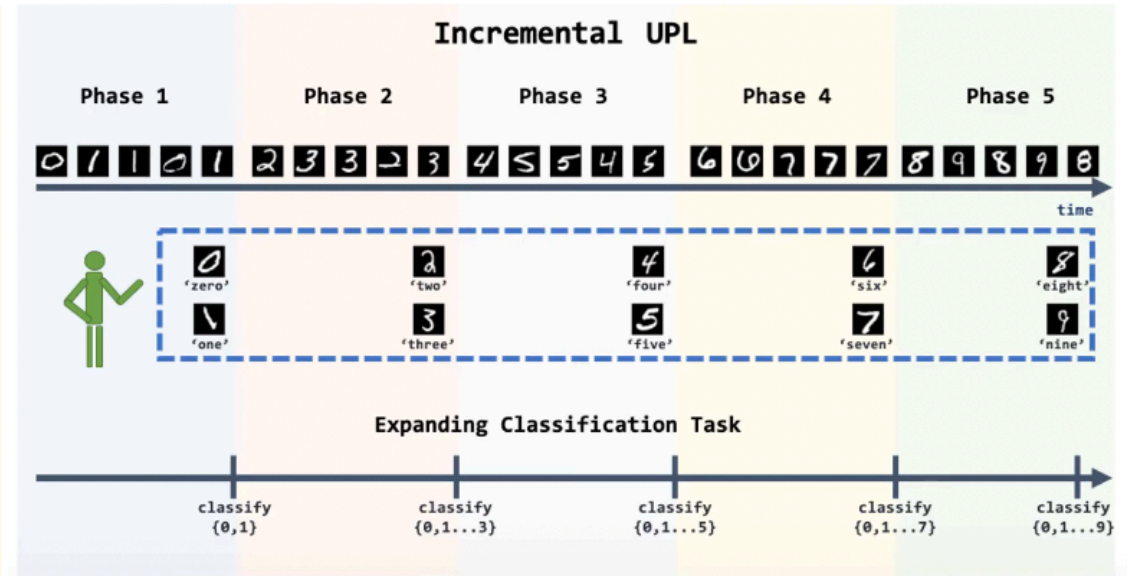
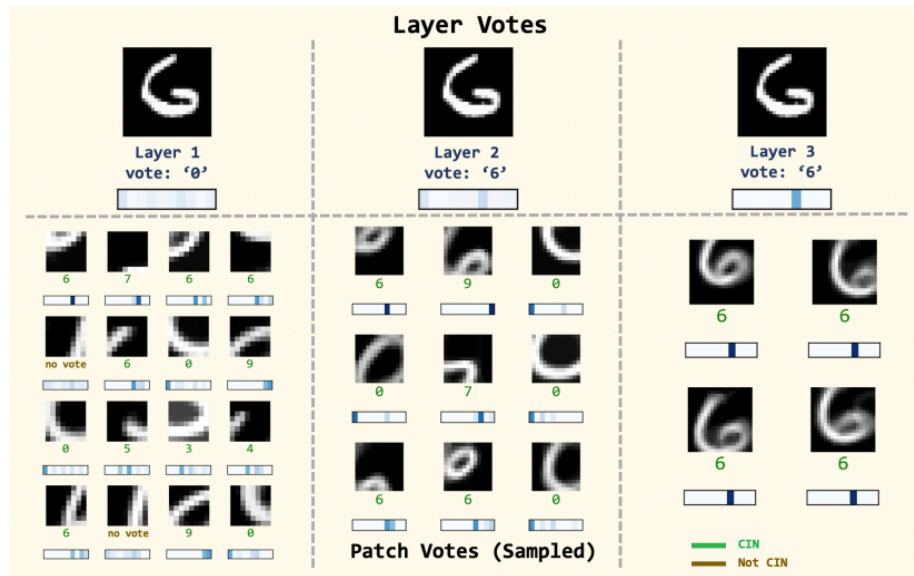
- Classification using STAM

: 소량의 labeled 데이터로 학습되었던 features를 각 클래스에 연관지어 성능 평가

- 하나의 이미지에 대한 모든 patch들은 가장 연관있다고 생각되는 CIN centroid에 대해서 투표를 하고, 최종 inference는 모든 layer에 거쳐서 가장 많은 누적 투표를 받은 클래스가 된다.

- Evaluation:

- 1) 클래스 분포가 각 phase마다 2개씩 증가하며 한번씩만 등장하는 데이터 스트림을 가정
- 2) 모든 클래스가 동일한 확률로 데이터 스트림에 등장



- 기존에 학습되었던 클래스에 대한 분류 성능의 유지 (left column)
- layer-2와 layer-3이 CIN centroids의 가장 높은 부분을 차지함 (center column)
- LTM centroid count이 새로운 클래스가 발생하기 전까지 stable. 즉, novelty detection이 가능함 (right column)

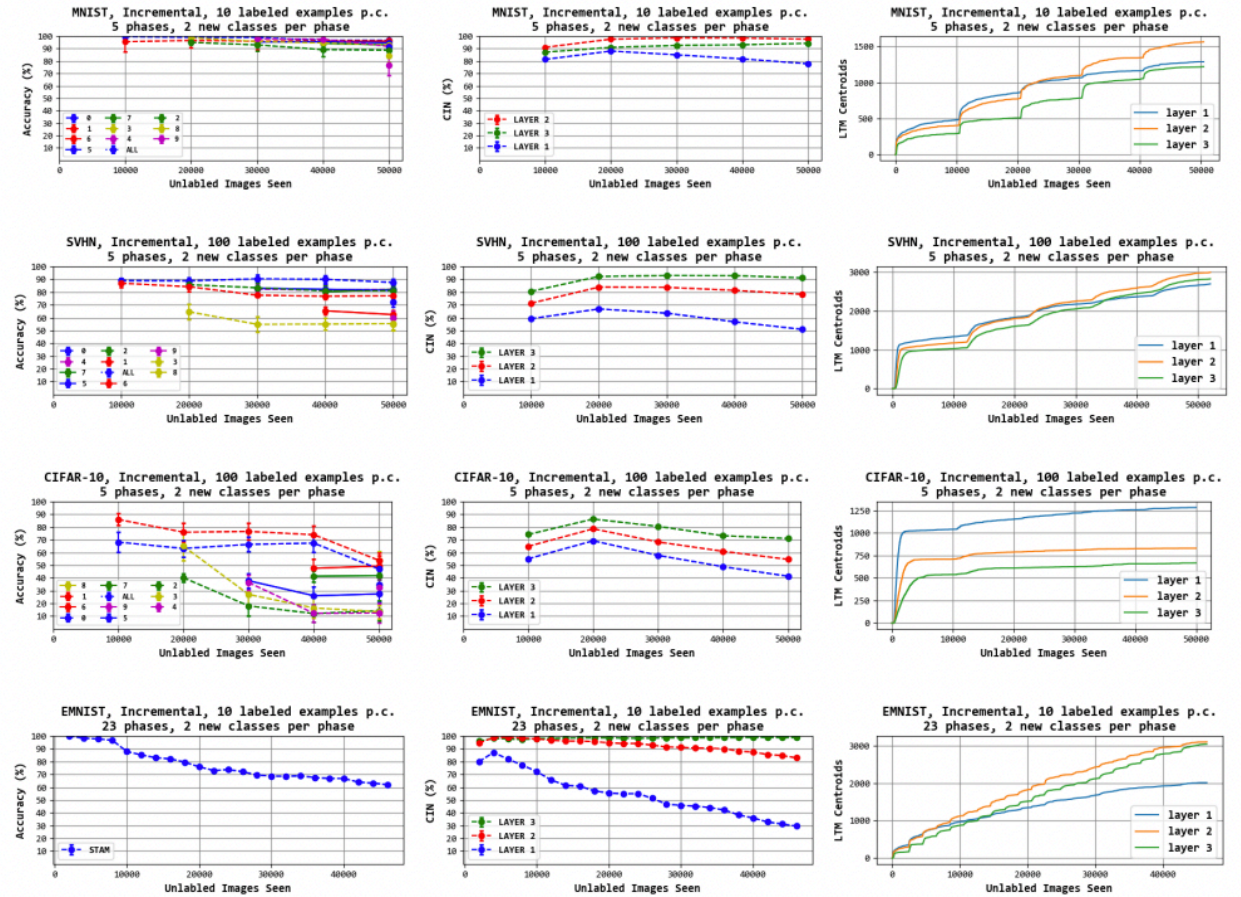


Figure 6: STAM Incremental UPL evaluation for MNIST (row-1), SVHN (row-2), EMNIST (row-3) and CIFAR-10 (row-4). Per-class (p.c.) and average classification accuracy (left); fraction of CIN centroids over time (center); number of LTM centroids over time (right). The task is expanding classification, i.e., recognize all classes seen so far.

2. Unsupervised Continual Learning in Streaming Environments

IEEE Transactions on Neural Networks and Learning Systems 2022

- 실시간 데이터 스트림에서 labeling하는 것은 많은 cost를 필요로 한다
따라서, 비지도 학습 기반의 DCN 구성 (DL과 Clustering을 동시에 수행)
- Autonomous Deep Learning (Dynamic network width & depth)
 - 1) FEL과 self-evolving FCN
 - 2) 각 FCN의 embedding space의 self-clustering
 - 3) latent-based regularization
- Fully Unsupervised Approach
 - : 학습은 라벨 없이 진행
 - 라벨은 class-associate representation의 추론을 위해서만 사용

- Feature Extractor는 clustering을 위한 representation의 학습을 위한 방식으로 업데이트
- FCN은 Stacked Auto Encoder의 역할을 대체하며, natural features(Z)를 clustering-friendly한 latent space로 매칭하는 역할
- SAE의 각 layer에서 self-clustering을 수행하여 나온 클러스터 allegiance값 (centroid값과 클러스터 개수를 이용해서 구함)과 cluster의 거리의 combined score에 기반하여 local score를 구함
- Classification 과정에서는 centroid를 각 클래스에 매칭하기 위하여 labeled 샘플들이 제공
- 모든 L 개의 layer에서의 local score의 합에 기반하여 predicted label이 결정됨

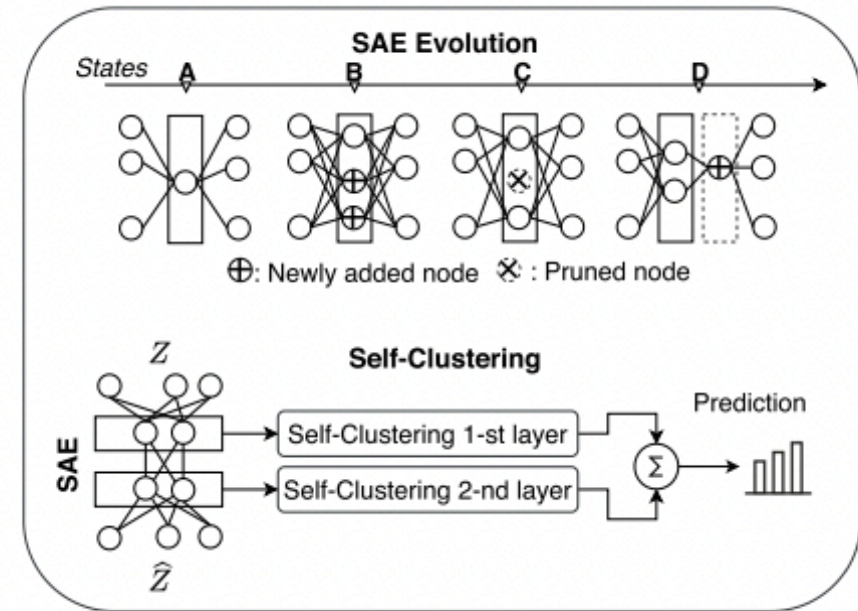


Fig. 1. ADCN network evolution. **State A**: SAE starts its learning process with a simple structure. **State B**: Two hidden nodes are added. **State C**: A hidden node is pruned. **State D**: A hidden layer is constructed. Each hidden layer is connected to a self-clustering mechanism.

- 실험 구성:
 - 1) 동일한 클래스의 unseen data가 다른 분포로 들어오는 환경
 - 2) unseen 클래스가 subsequent 배치에 들어오는 환경
 -> both are one-pass learning fashion
- STAM보다 떨어지는 경우도 존재하지만, network와 learning capacity를 필요에 따라 증가할 수 있다는 점에서 의미있음
- 라벨링된 샘플 200개 이상부터는 일정한 accuracy를 유지
 - > 모델 자체의 업데이트에는 라벨링된 샘플이 관여하지 않음

TABLE III
PERFORMANCE METRICS ON CONTINUAL LEARNING SCENARIO

Datasets	Methods	BWT	FWT	Task Acc. (%)	Preq. Acc. (%)
RMNIST	ADCN	-0.89 ± 1.1	41 \pm 1.35	79.64 \pm 0.68	78.41 \pm 0.46
	STAM	0.9 \pm 0.35	$30 \pm 0.37^\times$	$77.58 \pm 0.43^\times$	$74.71 \pm 0.29^\times$
	DCN+LwF	$-15 \pm 6.54^\times$	$16 \pm 5.50^\times$	$39.47 \pm 10.13^\times$	$52.79 \pm 13.54^\times$
	DCN+SI	$-13 \pm 6.00^\times$	$17 \pm 6.92^\times$	$44.62 \pm 12.46^\times$	$55.66 \pm 14.83^\times$
	AE+KM+LwF	$-18 \pm 2.32^\times$	$18 \pm 1.79^\times$	$45.31 \pm 1.63^\times$	$60.15 \pm 1.54^\times$
	AE+KM+SI	$-9 \pm 2.72^\times$	$16 \pm 2.15^\times$	$49.07 \pm 0.73^\times$	$51.19 \pm 1.39^\times$
PMNIST	ADCN	-20 ± 2.72	2 ± 2.23	21.18 ± 1.76	34.88 ± 1.84
	STAM	0.3 \pm 0.09	$1 \pm 0.44^\times$	47.97 \pm 0.59	55.37 \pm 0.26
	DCN+LwF	$-30 \pm 1.70^\times$	3 ± 1.42	35.53 ± 0.78	56.50 ± 0.54
	DCN+SI	$-43 \pm 3.23^\times$	1 ± 1.01	33.09 ± 2.14	64.87 ± 0.31
	AE+KM+LwF	$-28 \pm 1.71^\times$	3 ± 1.85	35.53 ± 1.02	56.27 ± 0.55
	AE+KM+SI	$-35 \pm 2.35^\times$	1 ± 1.97	36.06 ± 1.23	61.5 ± 0.36
SMNIST	ADCN	-9 ± 1.33	12 ± 2.53	83.40 ± 1.96	90.96 ± 0.59
	STAM	-2 \pm 0.13	0^\times	92.18 \pm 0.32	91.98 \pm 0.32
	DCN+LwF	-7 ± 3.98	18 ± 1.17	$52.42 \pm 5.02^\times$	$53.46 \pm 2.25^\times$
	DCN+SI	-4 ± 1.45	22 \pm 2.89	$58.82 \pm 1.18^\times$	$57.00 \pm 0.34^\times$
	AE+KM+LwF	-5 ± 1.11	18 ± 1.03	$55.12 \pm 0.97^\times$	$54.69 \pm 0.58^\times$
	AE+KM+SI	-3 ± 1.88	22 ± 1.73	$58.84 \pm 0.71^\times$	$56.58 \pm 0.28^\times$
SCIFAR10	ADCN	-22 ± 0.82	17 \pm 3.01	26.42 \pm 0.46	43.96 \pm 1.85
	STAM	-18 ± 2.46	0^\times	$20.6 \pm 0.66^\times$	$35.43 \pm 1.22^\times$
	DCN+LwF	-13 \pm 1.03	$11 \pm 1.12^\times$	$24.58 \pm 0.39^\times$	$32.67 \pm 0.34^\times$
	DCN+SI	-14 ± 0.72	$11 \pm 0.99^\times$	$24.36 \pm 0.74^\times$	$32.49 \pm 0.17^\times$
	AE+KM+LwF	-14 ± 0.79	$11 \pm 1.17^\times$	$24.00 \pm 0.47^\times$	$32.63 \pm 0.07^\times$
	AE+KM+SI	-14 ± 0.52	$10 \pm 0.51^\times$	$24.36 \pm 0.66^\times$	$32.25 \pm 0.27^\times$

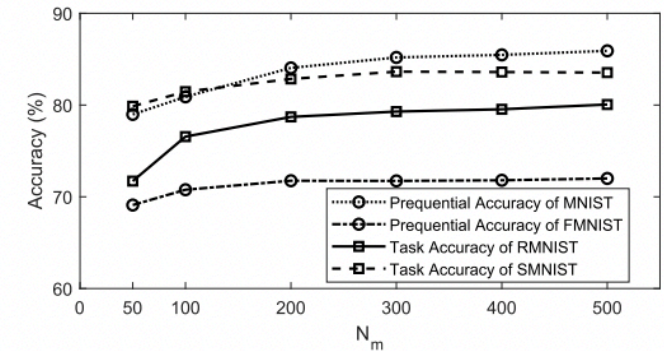


Fig. 2. Number of labeled samples per class N_m for associating centroids with classes is varied.

Impression

- NLP classification 문제에서 novelty detection 를 어떻게 정의할 것 인지 생각해보기
: NEI(Null Embedding Indicator) 혹은 Event Keyword값의 Embedding Vector 의 변화량(?)
- Centroid(clustering)와 Outlier에 대한 개념 적용
- Unsupervised한 manner로 New Instances와 New classes 각기 다른 스트림 환경에서의 실험

cf) Lbl2Vec

