Learning o Teach

Deep Learning Paper Review

한양대학교 AILAB 김 건

Introduction

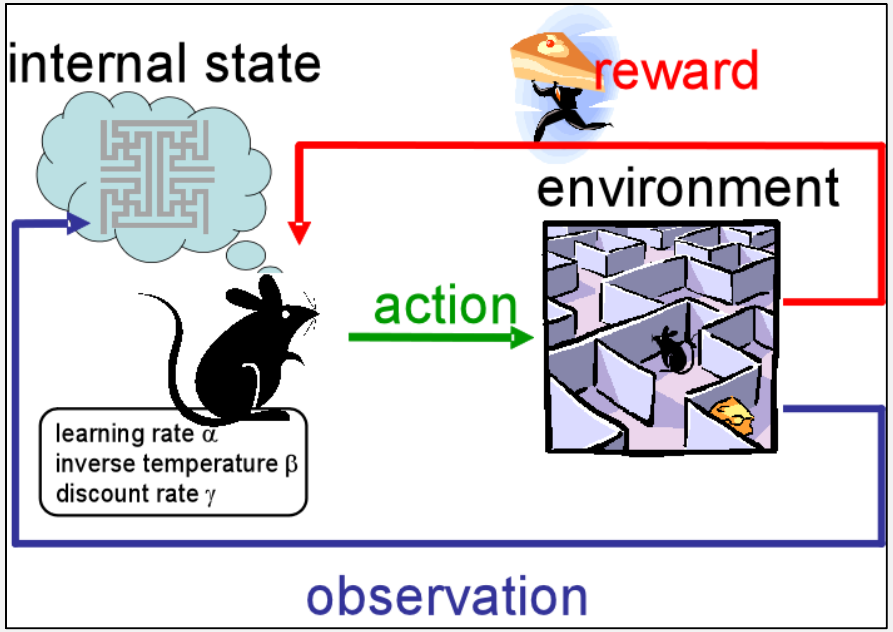
- Learning to Teach
- ICLR 2018 논문
- 최적화된 학습을 위해 자동으로 Data를 선별 학습시키는 프레임워크를 제안.
- 학습된 Teacher Model이 Data들을 필터링하여 Student에 제공, Student Agent들의 학습속도를 향상시킴

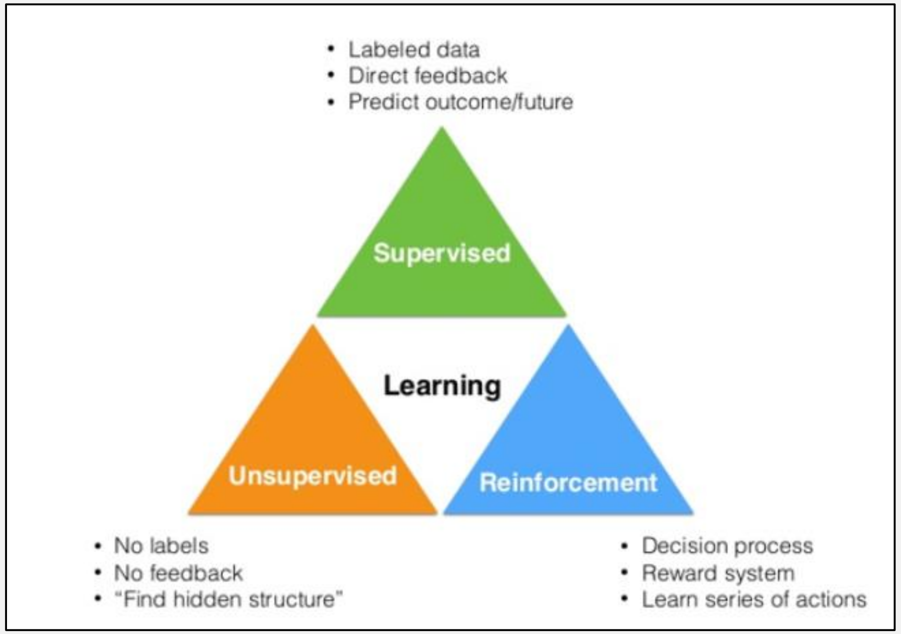
한양대학교

Introduction
Background
Model
Experiments
Conclusion

Reinforcement Learning

- Reinforcement Learning?
  - 어떤 환경(**Environment**) 안에서 정의된 에이전트(**Actor**)가 현재의 상태(**State**)를 인식하여, 선택 가능한 행동들(**Actions**) 중 보상(**Reward**)을 최대화 하는 행동 혹은 행동순서를 선택하는 방법이다. (Wikipedia)





- 자신의 행동에 대한 보상을 통해 행동들을 수정하는 학습
- 시행착오 학습

교육 손다이크가 발견한 학습 원리의 하나. 학습자가 목표에 도달하는 확실한 방법을 모르는 채 본능, 습관 따위에 의하여 시행과 착오를 되풀이하다가 우연히 성공한 동작을 계속함으로써 점차 시간을 절약하여 목표에 도달할 수 있게 된다는 원리이다. ≒시오법.

Q-Learning

- **Q-Learning 이란**
  - **Reinforcement Learning 가운데 가장 널리 사용되는 기계학습 알고리즘**
  - **현재 상태에서 선택 가능한 Action 중에 임의의 Action을 선택하고 실행한 뒤, 외부환경으로부터 Reward를 받는 과정을 반복함(시행착오)**
  - **Q function을 사용하여 Reward가 최대가 되는 행동 순서를 찾는 것을 목표로 함**

(1) State →

(2) Action →

Q

(3) Reward →

Q (state, action) = reward

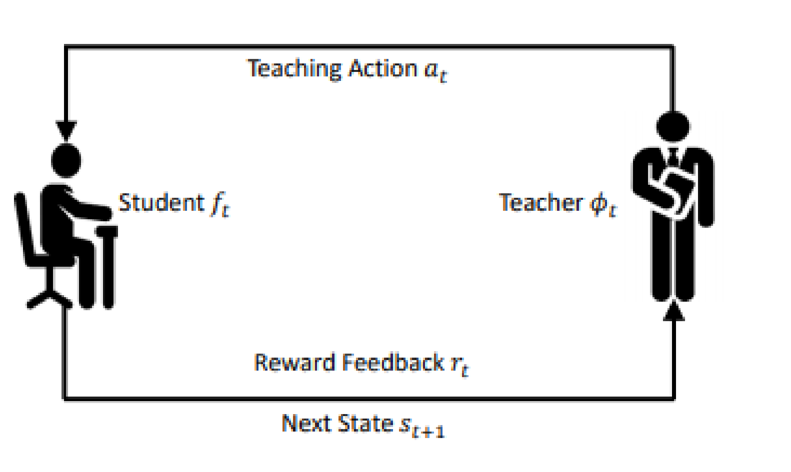
state와 action을 넣어주면 reward 값을 주는 함수

$$Q(s,a) \leftarrow r + \gamma \max_{a`} Q(s',a')$$

한양대학교

- Introduction
- Background
- Model
- Experiments
- Conclusion

L2T Framework



- Student Model(μ)
  - Input : Data, Loss Function, Function class
  - Output: w*로 최적화된 function
  - 기존 머신 러닝과 동일

$$\omega^* = \arg\min_{\omega \in \Omega} \sum_{(x,y) \in D} L(y, f_\omega(x)) \triangleq \mu(D, L, \Omega).$$

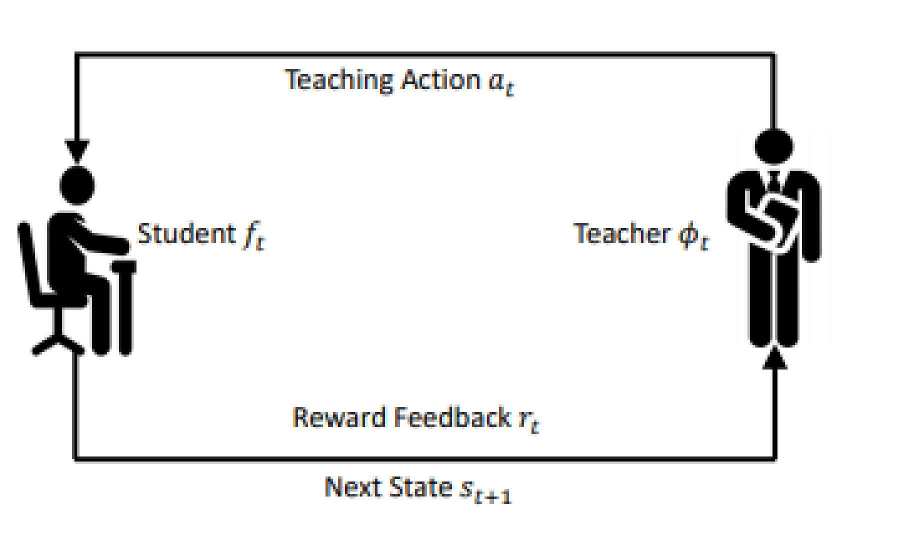
$$\min_{D,L,\Omega} \mathcal{M}(\mu(D, L, \Omega), D_{test}).$$

- Teacher Model (ϕ)
  - Regression을 빠르게 하는 최적의 D, L, Ω을 제공하는 모델
  - 논문에서는 L, Ω 학습 -> future work
  - 학습 속도, 성능을 높이는 Data를 필터링하는 Teacher Model을 제시 (L, Ω은 fix)
  - 논문에서는 3-Layer neural network (d * 12 * 1) 사용 d = g(s)의 차원

$$\max_\theta \sum_t r_t = \max_\theta \sum_t r(f_t) = \max_\theta \sum_t r(\mu(\phi_\theta(s_t), L, \Omega)),$$

L2T Framework

- Learning to Teach Framework



- Reinforcement feature
  - State Features g(s)

    Data features

    1) label category number

    Student model features

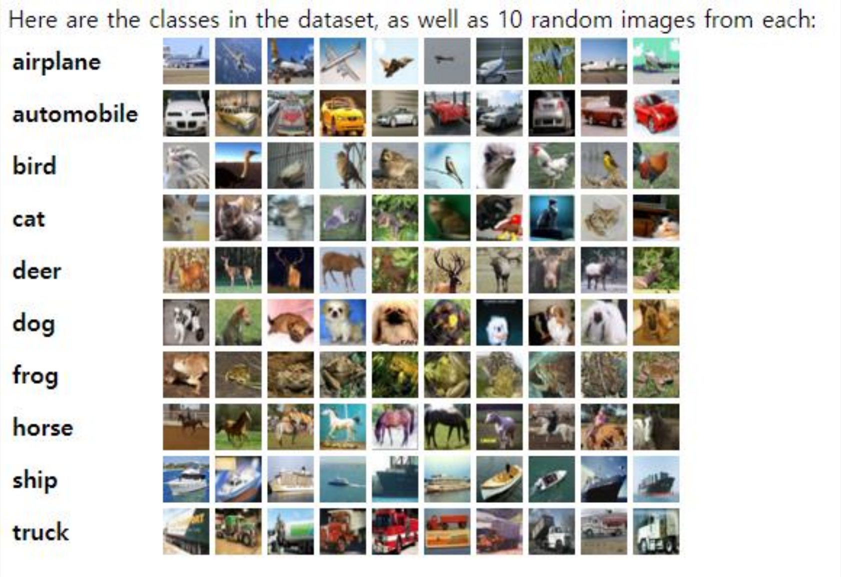
    1) iteration number
    2) average historical training loss
    3) Historical validation accuracy

    Combination features

    1) The predicted probabilities of each class
    2) The loss value on that data
    3) The margin value

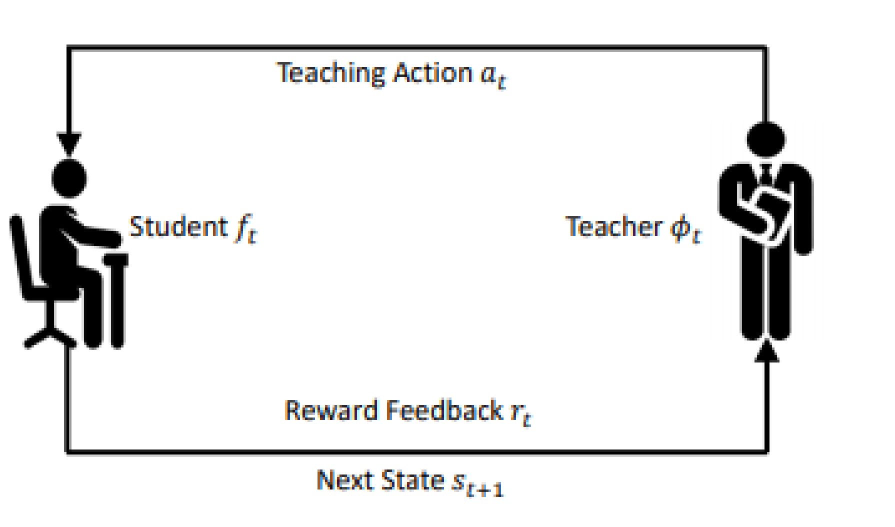
L2T Framework

- L2T Framework Setting Example CIFAR-10
  - ResNet32
  - Mini-batch size M = 128
  - Momentum-SGD



- g(s) 차원 = 10+3+12 = 25
  - Data features: 10 (라벨 수)
  - Model features: 3
  - Combined features: 10(클래스 별 확률) +2

L2T Framework



- Learning to Teach Framework

- Reinforcement feature
  - Reward $R_t = -\log(i_\tau / T`)$

    τ = accuracy threshold

    T` = maximum iteration number

    얼마나 빠르게 성능을 올렸는지를 Reward로 제공

  - Action a_t

    한 data instance의 training 참여 여부 결정 {0,1} = 데이터 필터링

  - Policy ϕ network (teacher model)

    Action 의 확률 결정

$$\phi_\theta(a|s) = a\sigma(w \cdot g(s) + b) + (1 - a)(1 - \sigma(\theta g(s) + b)).$$

한양대학교

- Introduction
- Background
- Model
- Experiments
- Conclusion

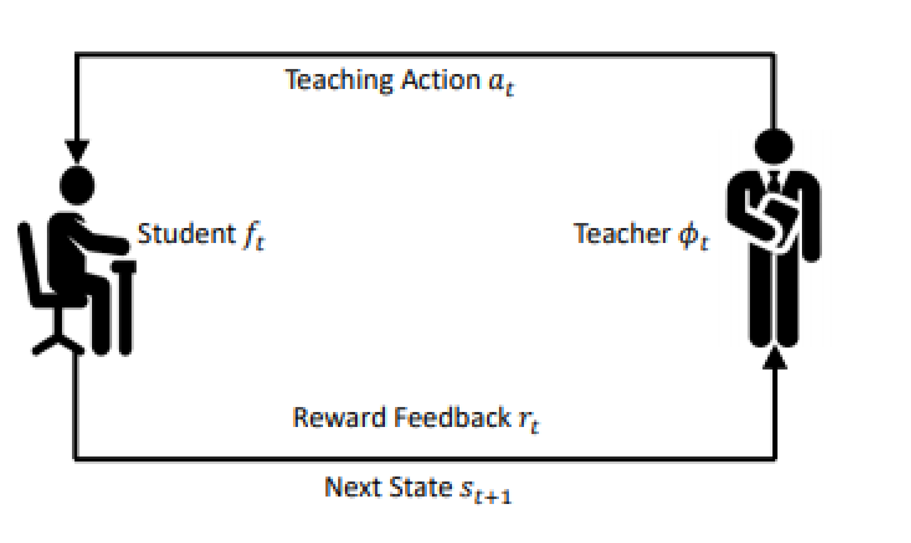
L2T Framework

- Learning to Teach Framework
- Student -> Teacher (ϕ)
  - Reward $R_t = -\log(i_{\tau}/T`)$

    τ = accuracy threshold

    T` = maximum iteration number

    얼마나 빠르게 성능을 올렸는지를 Reward로 제공

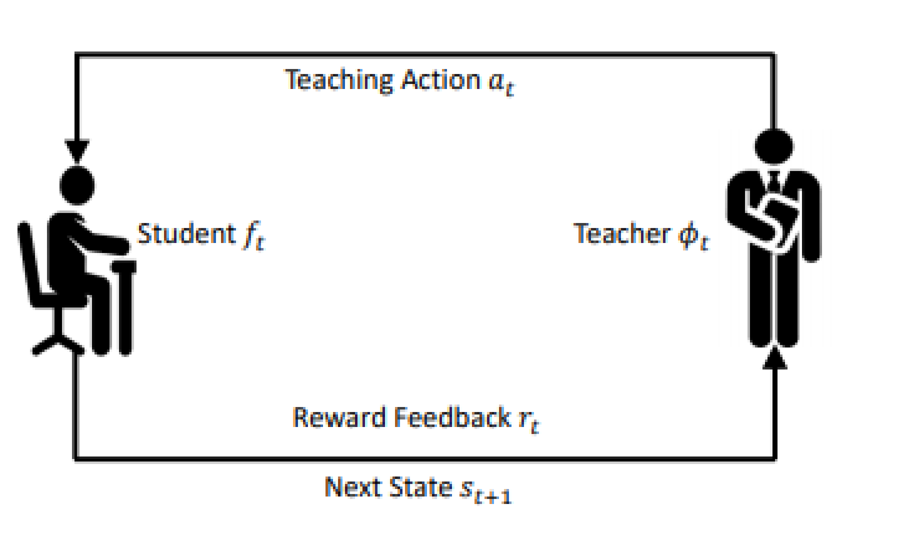


  - State Features g(s)의

- Student model features
  1) iteration number
  2) average historical training loss
  3) Historical validation accuracy
- Combination features
  1) The predicted probabilities of each class
  2) The loss value on that data
  3) The margin value

L2T Framework

- Learning to Teach Framework



- Teacher (ϕ) -> Student
  - Filtered data

- 정리하면 Teacher모델은 자기가 필터링한 데이터셋을 Student모델이 얼마나 빠르게 학습을 했는지를 토대로 필터링 함수를 학습함.

한양대학교

Introduction

Background

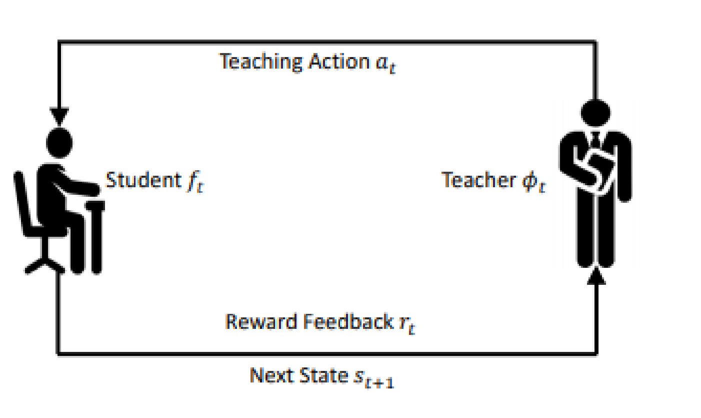
Model

Experiments

Conclusion

L2T Framework

- L2T Framework Example



Teacher Model 학습

1. Student가 첫번째 minibatch에 대하여 학습 진행
2. Data instance하나마다 g(s)가 계산이 됨.
3. Teacher Network에 g(s) 전달 -> 해당 loss를 학습에 반영시킬지 Teacher가 결정
4. 위 과정을 반복 수행 중에 Terminal State 도착하면 Reward를 Teacher 모델에 전달

   4-1) 정해진 accuracy threshold만큼 정확도가 상승한 경우

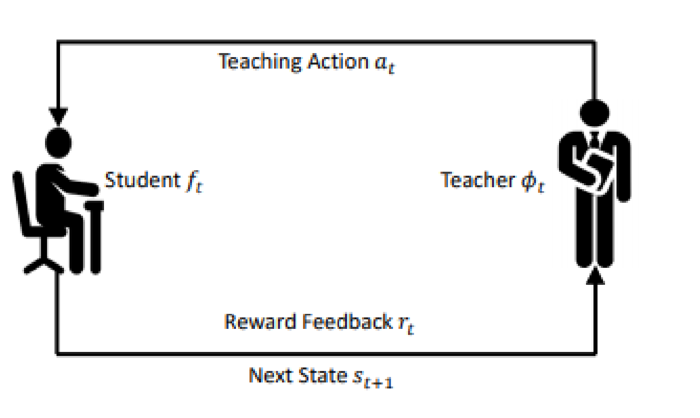
   4-2) 정해진 최대 iteration만큼 학습을 진행한 경우
5. Teacher Network를 policy gradient method로 학습

$$\nabla_\theta \approx \sum_{t=1}^{T} \nabla_\theta \log \phi_\theta(a_t|s_t) r_T.$$

각 state(feature)별 어떤 Action(filtering)을 하였을 때 기대 보상이 최대가 되는지 학습

L2T Framework

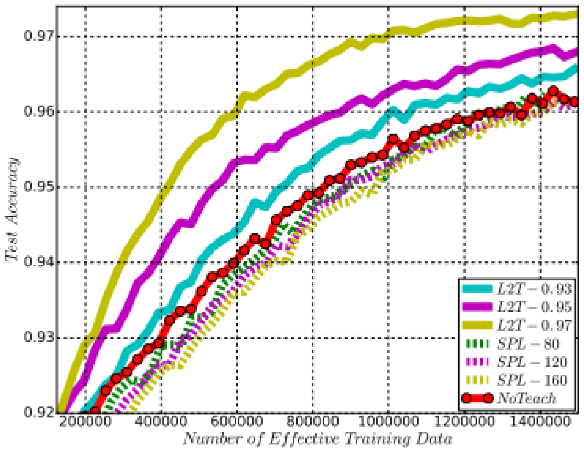
- L2T Framework Example



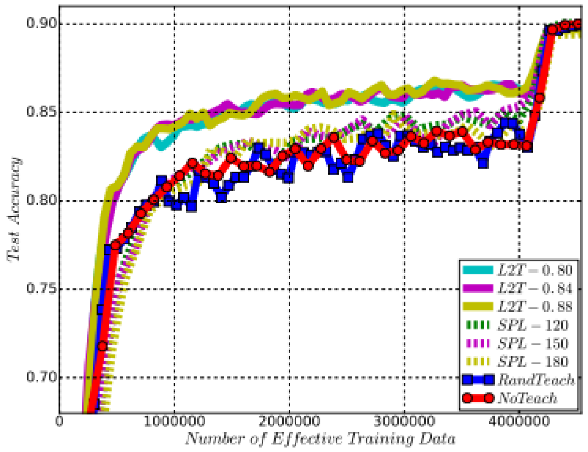
New Student 학습

1. 수렴한 teacher 모델
2. New Same Architecture Network 세팅(New Student)
3. Student 학습 진행
4. Data instance하나마다 g(s)가 계산이 됨.
5. Teacher Network에 g(s) 전달 -> 해당 loss를 학습에 반영시킬지 Teacher가 결정

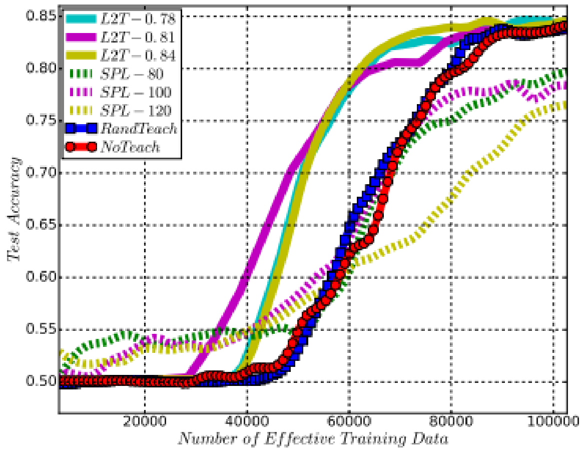
Experiments



(a) MNIST

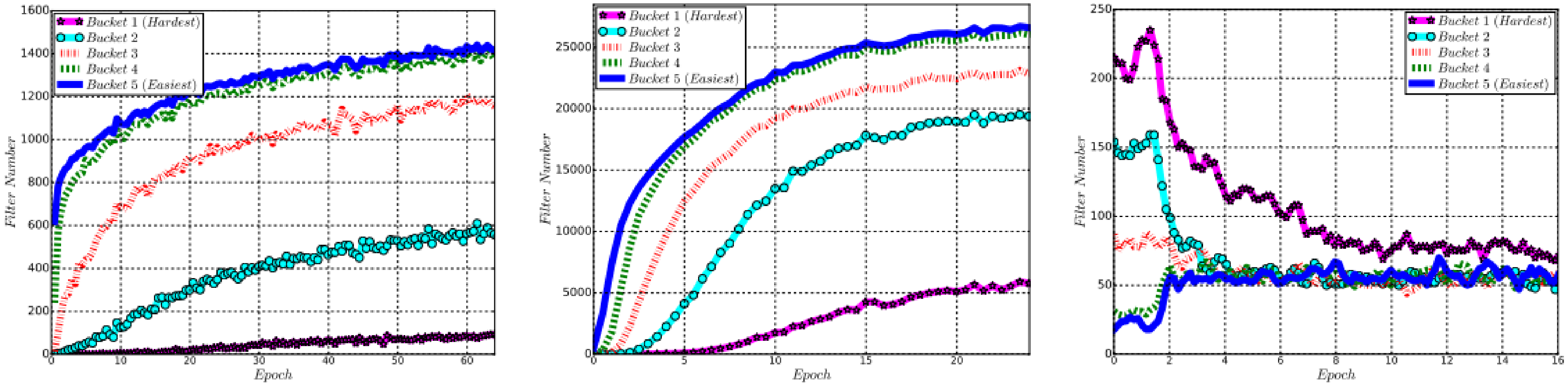


(b) CIFAR-10

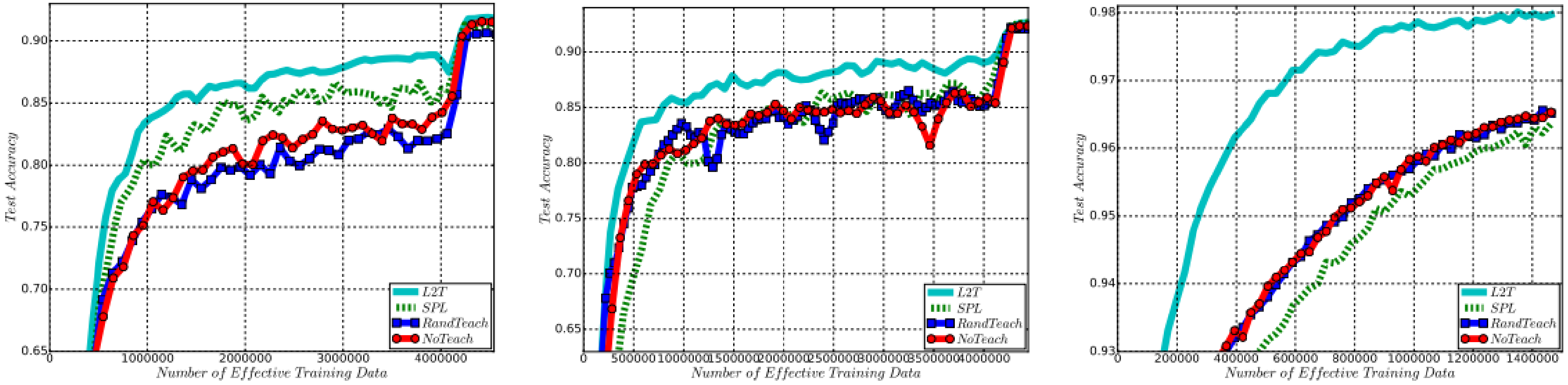


(c) IMDB

Experiments



(a) MNIST (b) CIFAR-10 (c) IMDB



(a) ResNet32 → ResNet110 (b) MNIST → CIFAR-10 (c) CIFAR10→MNIST

한양대학교

- Introduction
- Background
- Model
- Experiments
- Conclusion

Conclusion

- Conclusion

  - First, we have studied the application of L2T to image classification and sentiment analysis. We will study more applications such as machine translation and speech recognition.

  - we have focused on data teaching in this work. As stated in Subsection 3.1, we plan to investigate other teaching problems such as loss function teaching and hypothesis space teaching

  - we have empirically verified the L2T framework through experiments. It is interesting to build theoretical foundations for learning to teach, such as the consistence and generalization of the teacher model.

Thank you