

8. zero-shot, one-shot, few-shot

최근 나온 인공지능망 모델은 많은 학습 데이터를 요구한다.

많은 학습데이터를 제공함으로써 해당 문제에 대해 일반화를 잘 할 수 있게 되었지만, 데이터를 확보하는데 필요한 많은 시간과 돈이 필요하다.

모델이 적은 학습 데이터로(또는 그 문제에 대한 학습 데이터를 제공하지 않더라도) 일반화를 잘 할 수 있도록 관련 연구가 진행되고 있다.

특정 범주를 학습하는데 사용되는 데이터 수에 따라 zero-shot (가르치지 않은 범주를 일반화하는 기법), one-shot (범주 별로 한개의 학습 데이터 사용), few-shot (두개 이상, 일반적인 논문에서는 5개~6개 정도를 학습데이터로 사용)로 분류된다.

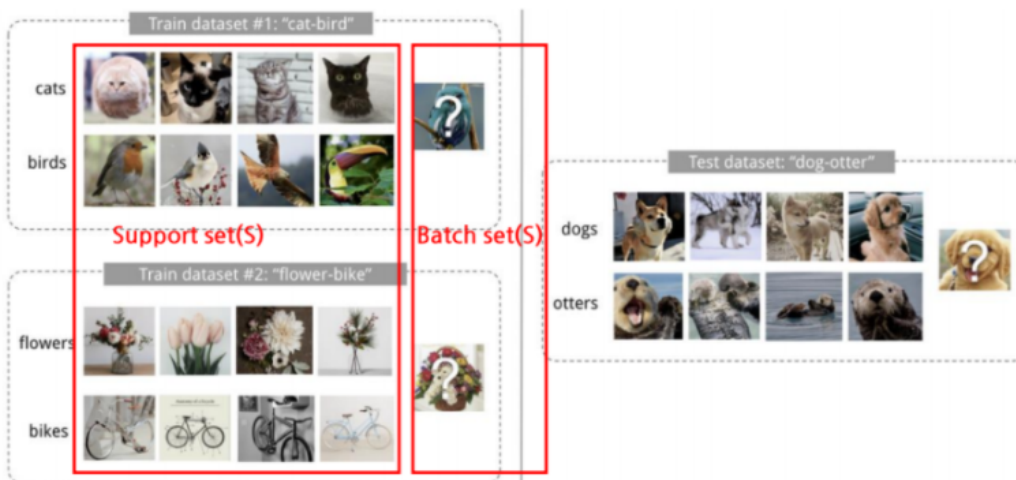
풀고자하는 문제

- M-way N-shot learning
 - M개의 class와 N개의 support data가 주어졌을 때 batch data(=query data)가 M개의 범주중 어디에 속하는가를 학습하는 문제
 - ex) 5개의 class가 주어지고 support data가 1개씩 주어지면, 5-way 1-shot learning
 - 일반적으로 사용되는 M과 N 값
 - M : 5 ~ 25
 - N : 1 ~ 5
 - 일반적인 문제해결 방식
 - Model-based
 - 학습한 Support data를 메모리에 저장한 뒤 batch data와 함께 모델 연산에 사용해 모델이 정답을 예측하는 기법
 - Metric-based (현 시점까지 리뷰한 논문들이 여기에 해당됨)
 - M*N개의 support data의 feature vector와 batch data를 거리계산 함수(코사인 유사도, 유클리드 거리)로 가장 근접한 class를 정답으로 선택
 - 모델은 다른 class의 feature vector를 잘 떨어뜨릴수있게 일반화하도록 학습
 - Optimization-based
 - M*N개의 support data로 모델이 학습을 해도 충분할수 있도록 모델의 파라미터 초기값을 최적화 하는 기법

Meta learning

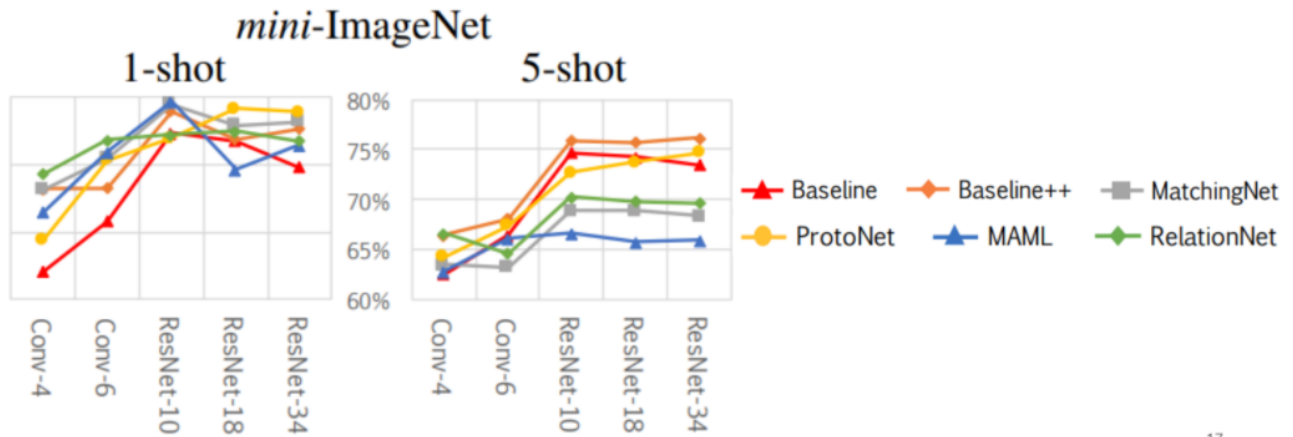
	Model-based	Metric-based	Optimization-based
Key idea	RNN; memory	Metric learning	Gradient Descent
How $P_{\theta}(y \mathbf{x})$ is modeled?	$f_{\theta}(\mathbf{x}, S)$	$\sum_{(\mathbf{x}_i, y_i) \in S} k_{\theta}(\mathbf{x}, \mathbf{x}_i) y_i (*)$	$P_{g_{\theta}(\theta, S^L)}(y \mathbf{x})$

- Episodic training
 - M-way N-shot learning의 테스트 데이터와 동일한 학습 데이터 구성을 맞추기 위해 제안한 기법
 - 데이터 구성
 - Episode 단위로 support data(M*N)와 batch data(B)가 주어짐
 - data는 supervised learning과 동일하게 (x,y) 구성
 - Loss 산출
 - metric based meta learning의 경우 batch data와 support data간의 거리차를 Loss로 사용



A CLOSER LOOK AT FEW-SHOT CLASSIFICATION (2019, ICLR)

Method	1-shot		5-shot	
	Reported	Ours	Reported	Ours
Baseline	-	42.11 \pm 0.71	-	62.53 \pm 0.69
Baseline* ³	41.08 \pm 0.70	36.35 \pm 0.64	51.04 \pm 0.65	54.50 \pm 0.66
MatchingNet ³ Vinyals et al. (2016)	43.56 \pm 0.84	48.14 \pm 0.78	55.31 \pm 0.73	63.48 \pm 0.66
ProtoNet	-	44.42 \pm 0.84	-	64.24 \pm 0.72
ProtoNet [#] Snell et al. (2017)	49.42 \pm 0.78	47.74 \pm 0.84	68.20 \pm 0.66	66.68 \pm 0.68
MAML Finn et al. (2017)	48.07 \pm 1.75	46.47 \pm 0.82	63.15 \pm 0.91	62.71 \pm 0.71
RelationNet Sung et al. (2018)	50.44 \pm 0.82	49.31 \pm 0.85	65.32 \pm 0.70	66.60 \pm 0.69



- 17 -

관련 자료

<https://github.com/sudharsan13296/Awesome-Meta-Learning>

<https://meta-learn.github.io/2019/>