Matching Networks for One Shot Learning

Author : Vinyals et al., 2016
21. 06.03 신한이

목차

Abstract

- 비전과 자연어 부문 연구에서 빠른 발전이 이루어짐.

- 하지만,

⇒ Standard supervised deep learning 분야에서 **적은 데이터**를 통해 새로운 컨셉을 학습하는 데에 **만족스러운 솔루션**이 없음.

⇒ 여전히 많은 데이터를 필요로 함

- 이를 해결하기 위해,

⇒ **External memory**를 추가한 NN 구조 도입

⇒ Small labelled support set만으로도 learning이 가능하다.

Related Works

Metric-Learning

- Matching Network
- Prototypical Network
- Relation Network

Embedding function 과
Distance 에 따라 달라짐.

- **Distance** : 임베딩 공간에서의 데이터 간 거리
- **Embedding func.** : 데이터를 저차원으로 임베딩

Introduction

Matching Network

Matching Network 에서는

convolution과 **lstm 기반 embedding function**과
cosine similarity distance를 사용했고

Meta-train set에서 **task sampling**을 통해 support, query set을 뽑은 후 query set을 잘 분류하기 위한 episode training을 통해 모델을 학습

Introduction

Few Shot Learning

[Few Shot Learning]

: 소수 데이터(support data) 로 다수 데이터(query data)를 예측해야 하는 과업에 적용되는 학습 기법

[One Shot Learning]

: which consists of learning a class from a single labelled example.

Introduction

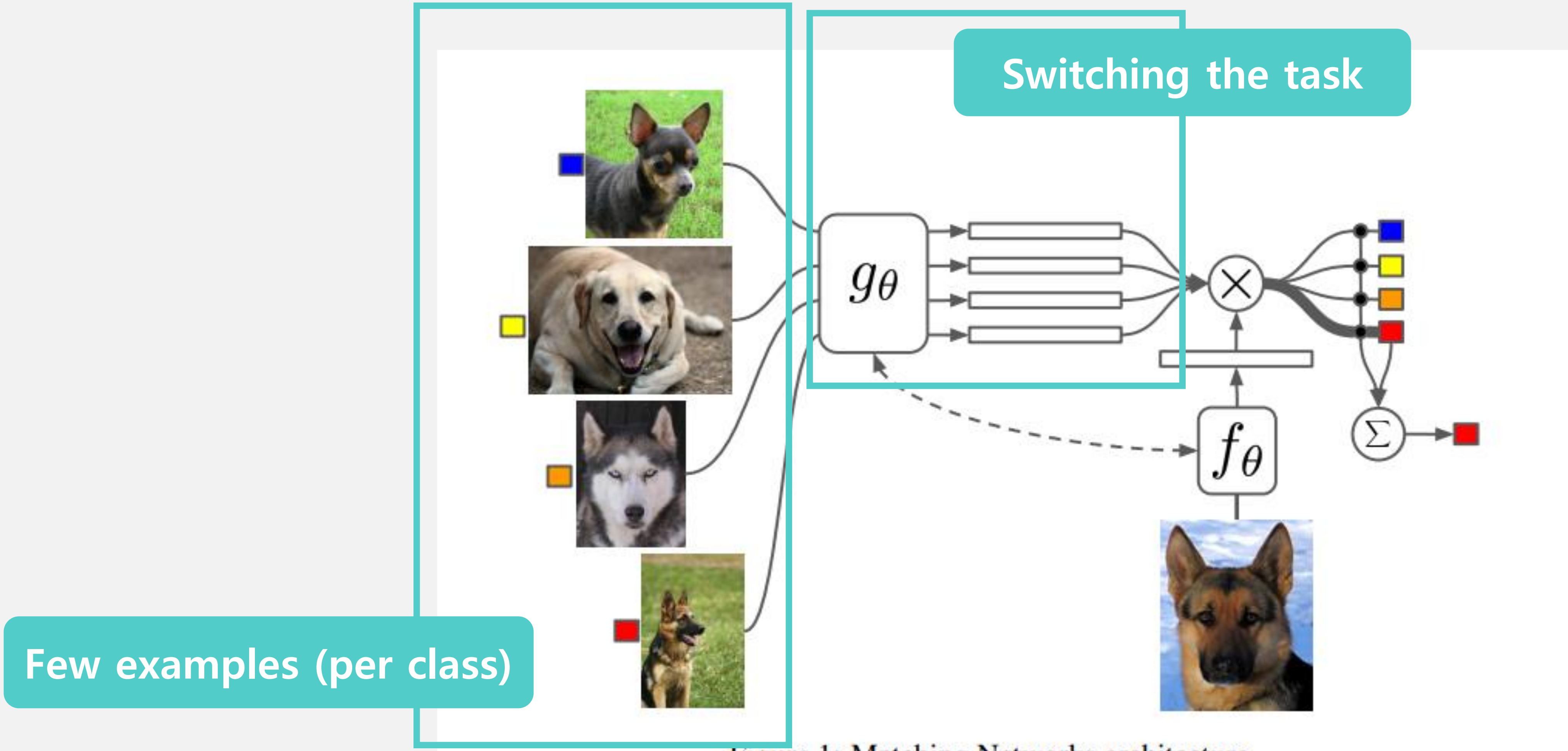

Figure 1: Matching Networks architecture

Method

[from Paper]

Our non-parametric approach to solving one-shot learning is based on **two components**.

First, our model architecture follows recent advances in neural networks **(1)augmented with memory** (as discussed in Section 3).

Second, we employ a **(2)training strategy** which is tailored for one-shot learning from the support set S.

Method

Model Architecture

Model Architecture : External memory 를 추가한 NN 구조 (cf. seq2seq, pointer NN)

Attention Mechanism

$$\hat{y} = \sum_{i=1}^{k} a(\hat{x}, x_i) y_i$$

Sample, from Support Set **S**

Method

Model Architecture

The Attention Kernel : a(^x, xi)를 정의

C : cosine distance

$$a(\hat{x}, x_i) = e^{c(f(\hat{x}), g(x_i))} / \sum_{j=1}^{k} e^{c(f(\hat{x}), g(x_j))}$$

f(x), g(x) : embedding func.

Method

Model Architecture

Full Context Embeddings : ^x를 임베딩하는 f도 S를 고려할 수 있다.

$$f(\hat{x}, S) = \text{attLSTM}(f'(\hat{x}), g(S), K)$$

- LSTM 구조를 사용
- 매 time step마다 같은 f값을 넣어준다.

Method

Training Strategy

Objective function 최대화

$$\theta = \arg\max_{\theta} E_{L\sim T}\left[E_{S\sim L,B\sim L}\left[\sum_{(x,y)\in B}\log P_{\theta}\left(y|x,S\right)\right]\right].$$

1) L을 T로부터 샘플링한다.
2) L을 이용해 support set S와 batch B를 샘플링한다.
3) Support set S를 이용해 B의 샘플에 대해 label을 예측한다.

Experiments

Model	Matching Fn	Fine Tune	5-way Acc		20-way Acc	
			1-shot	5-shot	1-shot	5-shot
PIXELS	Cosine	N	41.7%	63.2%	26.7%	42.6%
BASELINE CLASSIFIER	Cosine	N	80.0%	95.0%	69.5%	89.1%
BASELINE CLASSIFIER	Cosine	Y	82.3%	98.4%	70.6%	92.0%
BASELINE CLASSIFIER	Softmax	Y	86.0%	97.6%	72.9%	92.3%
MANN (NO CONV) [21]	Cosine	N	82.8%	94.9%	–	–
CONVOLUTIONAL SIAMESE NET [11]	Cosine	N	96.7%	98.4%	88.0%	96.5%
CONVOLUTIONAL SIAMESE NET [11]	Cosine	Y	97.3%	98.4%	88.1%	97.0%
MATCHING NETS (OURS)	Cosine	N	**98.1%**	**98.9%**	**93.8%**	98.5%
MATCHING NETS (OURS)	Cosine	Y	97.9%	98.7%	93.5%	**98.7%**

Table 1: Results on the Omniglot dataset.

=> 모두 N-way k-shot learning 작업

Conclusion

(* Metric-based approach의)

장점

- Metric 기반의 Meta learning은 단순하고 직관적이며, Embedding 의 시각화와 해석하기에 쉽다

단점

- domain adaptation에 약하다.
  (= domain간 차이가 큰 경우 adaptation이 어렵다.)
- 계산량이 많다.

참조 자료

+ Paper

  https://arxiv.org/pdf/1606.04080.pdf

- Refer

  https://jiminsun.github.io/2019-02-20/Vinyals-2016/

- Refer

  http://www.navisphere.net/6014/matching-networks-for-one-shot-learning/