1. Siamese Neural Networks for One-shot Image Recognition

논문 : https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

요약

학습은 값비싼 계산이 필요하고, 사용가능한 데이터가 적은 경우 학습이 어려움

이를 해결하는 대표적인 예로 새 범주당 하나의 데이터를 사용하는 one-shot learning이 있음

본 논문은 입력간의 유사성을 Siamese Neural Networks가 학습할 수 있도록 제안함 (Contrastive Learning과 유사)

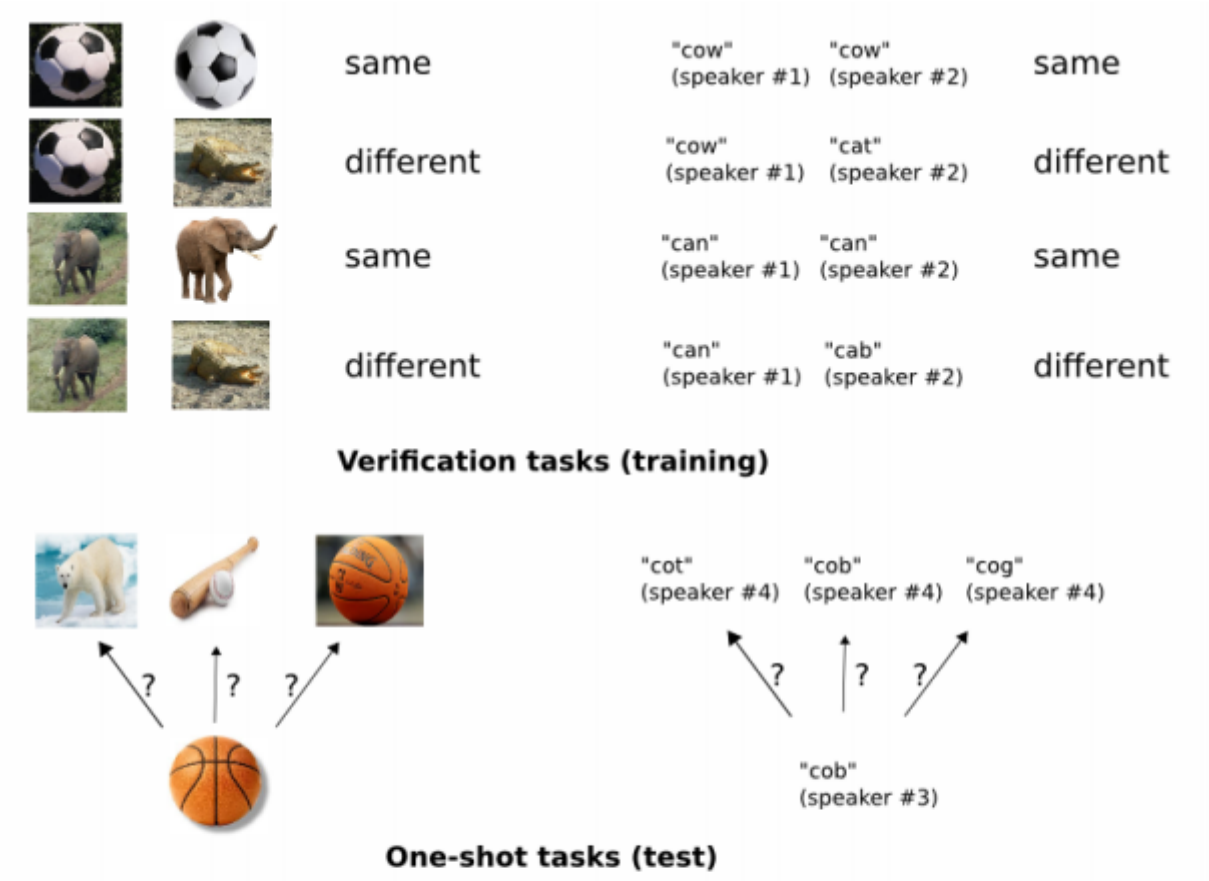

Figure 2. Our general strategy. 1) Train a model to discriminate between a collection of same/different pairs. 2) Generalize to evaluate new categories based on learned feature mappings for verification.

모델 구조

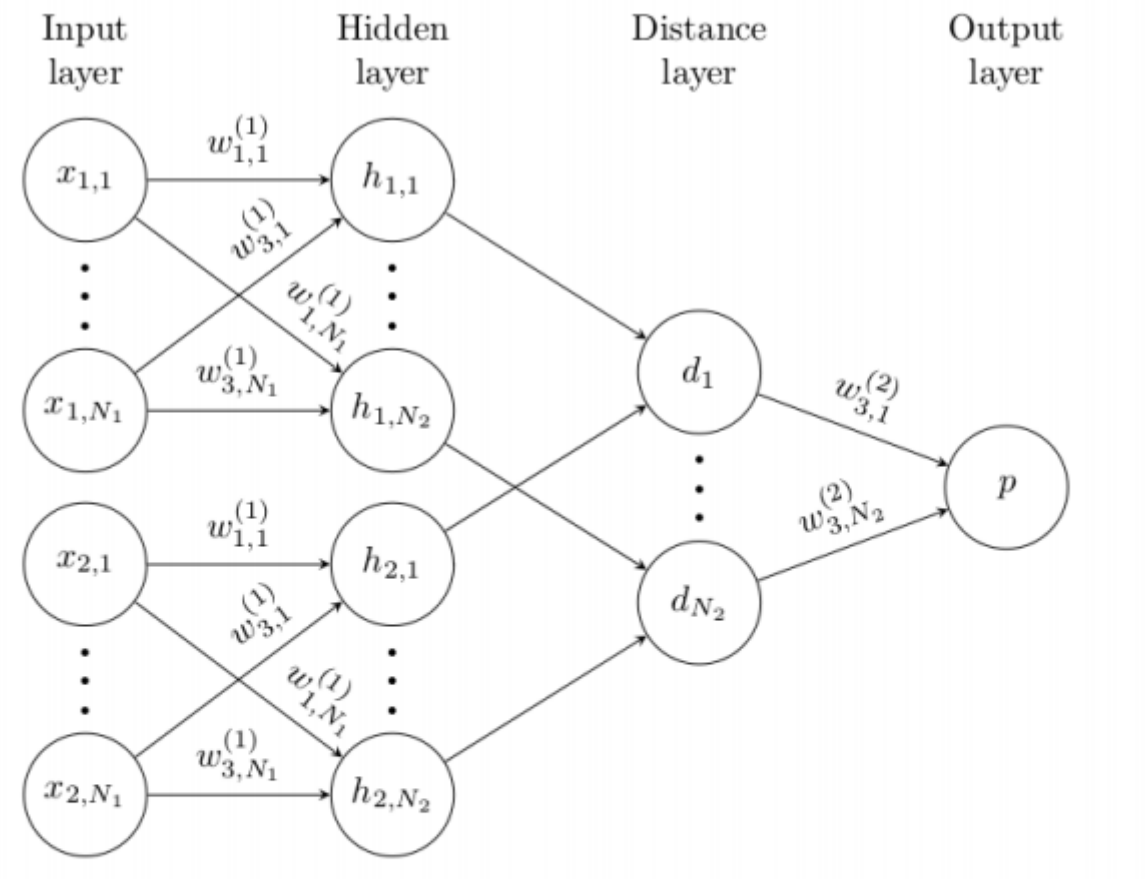

Figure 3. A simple 2 hidden layer siamese network for binary classification with logistic prediction p. The structure of the network is replicated across the top and bottom sections to form twin networks, with shared weight matrices at each layer.

- Input Layer 부터 Hidden Layer 까지 모델의 동일한 파라미터를 사용

$\mathbf{h}^{(j)}_{1,L-1}$: 첫번째 비교 이미지의 Feature Vector의 j번째 차원 값, $\mathbf{h}^{(j)}_{2,L-1}$: 두번째 비교 이미지의 Feature Vector의 j번째 차원 값

- 해당 모델은 두 데이터 간 유사도를 예측함 (sigmoid 사용)
- 모델은 같은 범주의 입력의 쌍인 경우 1, 다른 범주의 입력의 쌍인 경우 0에 가깝게 예측

$$\mathbf{p} = \sigma\left(\sum_j \alpha_j \left|\mathbf{h}^{(j)}_{1,L-1} - \mathbf{h}^{(j)}_{2,L-1}\right|\right)$$

σ : Sigmoid , α_j : 학습 가능한 파라미터, $\left|\mathbf{h}^{(j)}_{1,L-1} - \mathbf{h}^{(j)}_{2,L-1}\right|$: 그림 3에서 d 로 표현

- 테스트 셋이 입력되면 다른 클래스들과 대조해서 가장 높은 p를 가진 범주로 예측
- Hidden Layer의 경우 Omniglot dataset의 모든 데이터를 사용했을 때의 가장 성능이 좋은 모델 구조를 사용 (그림 4)

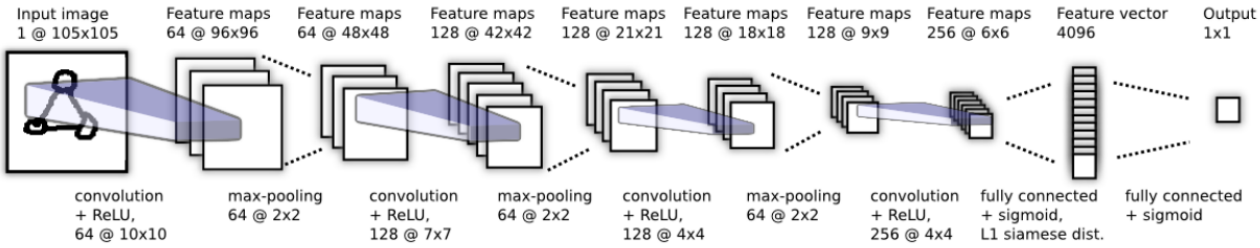

Figure 4. Best convolutional architecture selected for verification task. Siamese twin is not depicted, but joins immediately after the 4096 unit fully-connected layer where the L1 component-wise distance between vectors is computed.

Loss Function

- 일반적인 BinaryCrossEntropy Loss

$$\mathcal{L}(x_1^{(i)}, x_2^{(i)}) = \mathbf{y}(x_1^{(i)}, x_2^{(i)}) \log \mathbf{p}(x_1^{(i)}, x_2^{(i)}) + (1 - \mathbf{y}(x_1^{(i)}, x_2^{(i)})) \log (1 - \mathbf{p}(x_1^{(i)}, x_2^{(i)})) + \boldsymbol{\lambda}^T |\mathbf{w}|^2$$

실험

- 데이터 셋 : Omniglot Dataset
  - 여러 나라의 50 종류 문자 이미지. (영어, 라틴어, 한글)
  - 다른 모델과 성능 비교를 위해 40 종류의 문자를 모델 튜닝에 사용 / 10 종류의 문자를 one-shot learning 성능 비교에 사용
- 실험 방법
  - Omniglot Dataset을 학습에 다 사용했을 때의 정확도 (표 1)
    - 최고의 모델 구조와 데이터 확장기법의 유효성을 검증하기 위해 실험
    - 이미지의 쌍을 3만, 9만, 15만개 생성해서 학습에 사용했을 때 정확도 산출
    - 데이터 확장은 Affine transformation을 사용

      Affine distortions. In addition, we augmented the training set with small affine distortions (Figure 5). For each image pair $\mathbf{x_1}, \mathbf{x_2}$, we generated a pair of affine transformations T_1, T_2 to yield $\mathbf{x'_1} = T_1(\mathbf{x_1}), \mathbf{x'_2} = T_2(\mathbf{x_2})$, where T_1, T_2 are determined stochastically by a multi-dimensional uniform distribution. So for an arbitrary transform T, we have $T = (\theta, \rho_x, \rho_y, s_x, s_y, t_x, t_x)$, with $\theta \in [-10.0, 10.0], \rho_x, \rho_y \in [-0.3, 0.3], s_x, s_y \in [0.8, 1.2]$, and $t_x, t_y \in [-2, 2]$. Each of these components of the transformation is included with probability 0.5.
  - Omniglot Dataset의 one-shot learning (표 2)
    - 한 개의 범주에 한 개의 학습데이터를 사용하여 다른 기법들과 성능 비교
  - MNIST Dataset으로 one-shot learning (표 3)
    - MNIST Dataset에 one-shot learning 실험

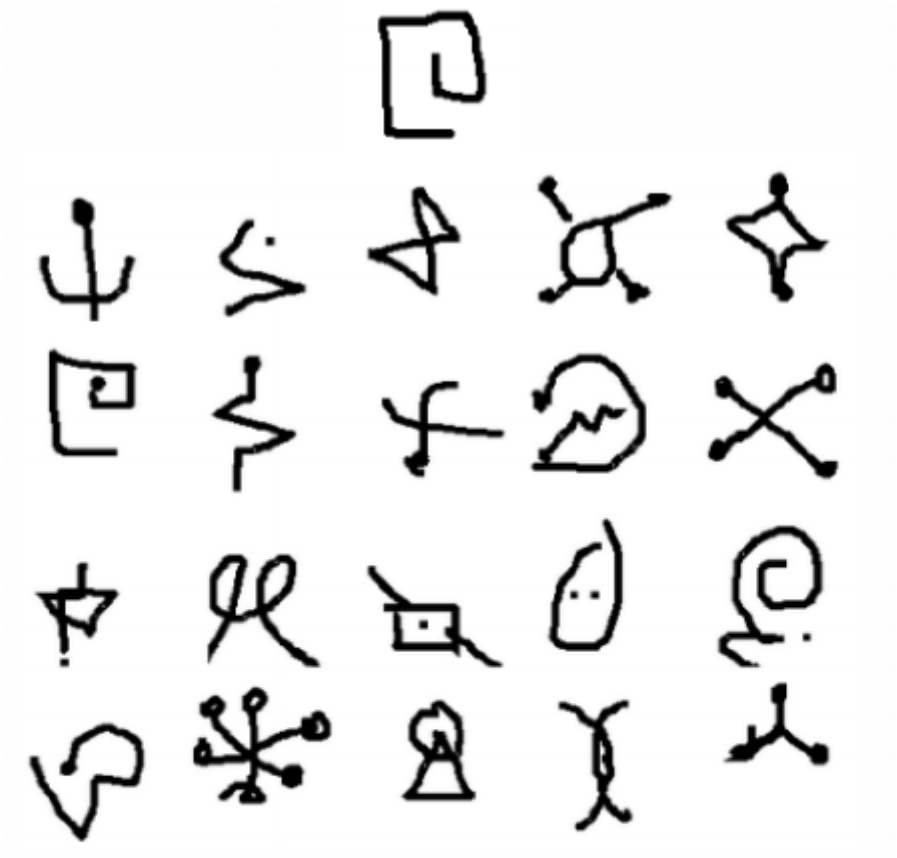

Figure 1. Example of a 20-way one-shot classification task using the Omniglot dataset. The lone test image is shown above the grid of 20 images representing the possible unseen classes that we can choose for the test image. These 20 images are our only known examples of each of those classes.

Table 1. Accuracy on Omniglot verification task (siamese convolutional neural net)

Method	Test
30k training	
no distortions	90.61
affine distortions x8	91.90
90k training	
no distortions	91.54
affine distortions x8	93.15
150k training	
no distortions	91.63
affine distortions x8	**93.42**

Table 2. Comparing best one-shot accuracy from each type of network against baselines.

Method	Test
Humans	95.5
Hierarchical Bayesian Program Learning	95.2
Affine model	81.8
Hierarchical Deep	65.2
Deep Boltzmann Machine	62.0
Simple Stroke	35.2
1-Nearest Neighbor	21.7
Siamese Neural Net	58.3
Convolutional Siamese Net	92.0

Table 3. Results from MNIST 10-versus-1 one-shot classification task.

Method	Test
1-Nearest Neighbor	26.5
Convolutional Siamese Net	70.3