



期末報告

Siamese Neural Networks for One-shot Image Recognition



Contents

- 01. Introduction
- 02. Proposed Method
- 03. Experimental Results
- 04. Conclusion



Introduction

One-shot Learning

一次性學習

面對分類問題時，在預測測試實例的類別之前，對每一類別只提供一個或者少量的訓練樣本，利用輸入間的相似度進行分類。

處理方法

每一類別需發展出主要特定的特徵、對目標類別的推論過程中具高度判別特性

衍生問題

發展出的分類系統在相似的實例下會有好的分類結果；
但在對沒有被標示類別的新數據進行預測時，演算法容易崩潰。

目標

- > 模型用於不熟悉的類別，不需再重新進行訓練
- > 經過訓練後，可以成功判別新數據，預測其所屬未知分佈的類別。

Proposed Method

Siamese Neural Network—架構

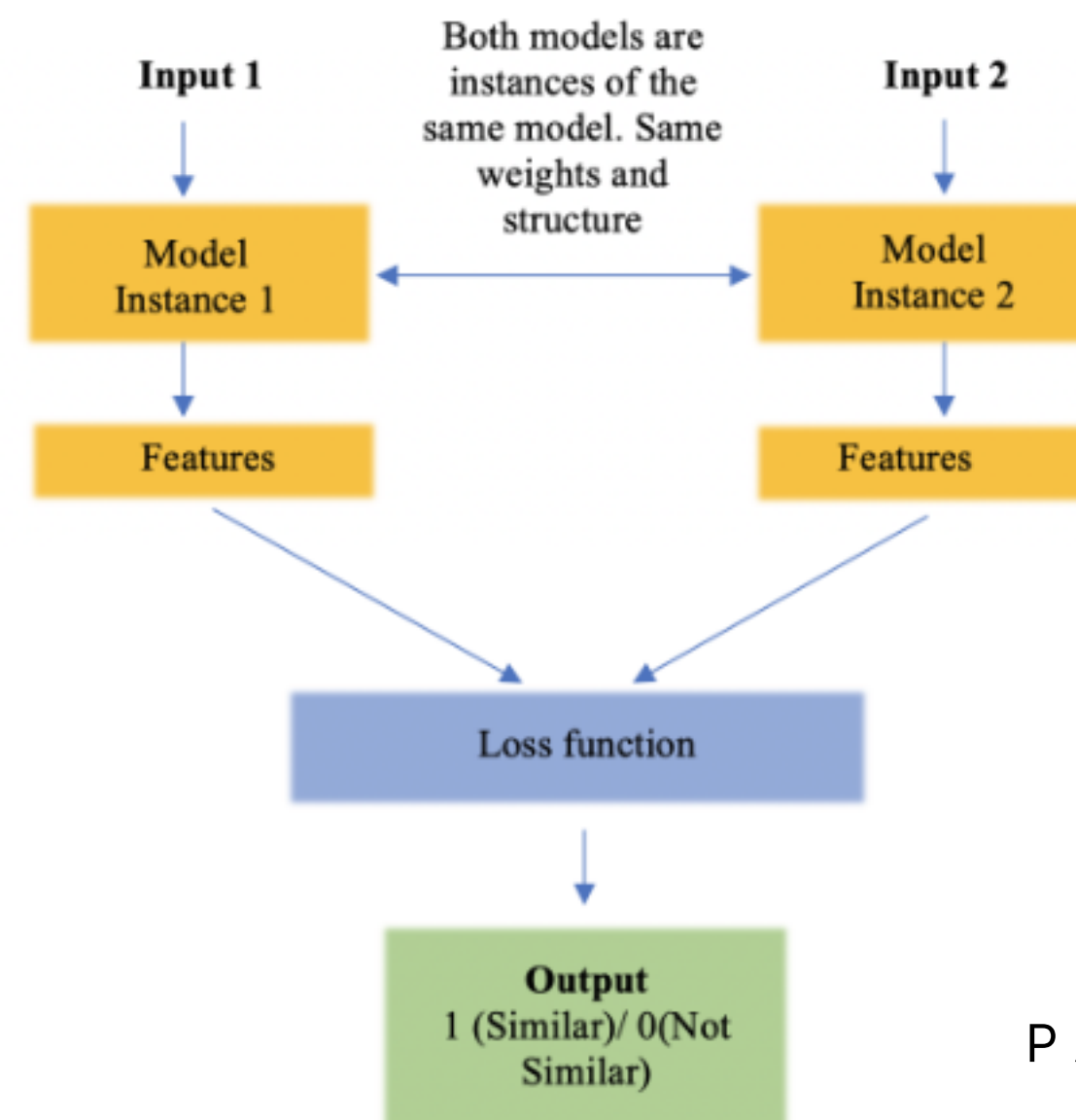
》 使用大量卷積層

- > 能夠提取一般圖片的特徵，且有效預測未知分佈的類別。
- > 容易利用配對的資料來訓練。
- > 不依賴於特定領域的知識，而是利用深度學習的技術。

Proposed Method

Siamese Neural Network—架構

- > 透過限制對輸入結構的假設，同時獲取特徵，使模型能夠從幾個範例中成功泛化。
- > 由兩個可接受不同輸入的神經網路組成，透過 loss function 連接。
- > 計算輸入之間的相似性進行分類預測。
- > 孿生網路之間的參數是綁定且對稱，確保兩個相似的圖像不會因各自的網路映射到特徵空間中完全不同的位置。







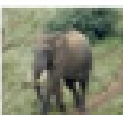



Proposed Method

Siamese Neural Network—步驟

》開發一次性圖像分類的模型

1. 建立一能夠區分圖像對類別標識的神經網路。
2. 假設此神經網路能夠泛化到一次性分類。
3. 將測試圖像和新圖像配對後進行評估，找出得到驗證網路最高分的配對，即為最可能屬於同一類別的圖像。

		same	"cow" (speaker #1)	"cow" (speaker #2)	same
		different	"cow" (speaker #1)	"cat" (speaker #2)	different
		same	"can" (speaker #1)	"can" (speaker #2)	same
		different	"can" (speaker #1)	"cab" (speaker #2)	different

Verification tasks (training)



One-shot tasks (test)

Experimental Results

dataset：全語言文字數據集，含有50種語言的字母表，每個字符有 20 個不同的人書寫

> 訓練驗證網路

》總數據的 60% 用於訓練：50 個語言中的 30 個語言和 20 個書寫者中抽 12 個。

》隨機抽取配對後資料，測試三種不同大小的數據集，分別為 30,000、90,000 和 150,000 個訓練樣本。

> 全域仿射

在每一數據集中增加 8 個轉換

ちちちちち	世世世世世	ぬぬぬぬぬ
ちちちちち	世世世世世	ぬぬぬぬぬ
ちちちちち	世世世世世	ぬぬぬぬぬ
ちちちちち	世世世世世	ぬぬぬぬぬ
ちちちちち	世世世世世	ぬぬぬぬぬ

Experimental Results

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

Method	Test
30k training	
<i>no distortions</i>	90.61
<i>affine distortions x8</i>	91.90
	270k training
90k training	
<i>no distortions</i>	91.54
<i>affine distortions x8</i>	93.15
	810k training
150k training	
<i>no distortions</i>	91.63
<i>affine distortions x8</i>	93.42
	1350k training

準確率最高

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

在現有一次性分類模型中，
展示出不錯的預測結果。

Experimental Results

evaluate dataset : 手寫辨識資料集 (使用全語言文字數據集進行訓練)

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

Method	Test
1-Nearest Neighbor	26.5
Convolutional Siamese Net	70.3

以一般化模型而言
結果可被接受

Conclusion

結論：

> 在現有一次性分類模型中，展示出不錯的預測結果。

未來展望：

> 目前透過全域仿射變換處理圖像對，該演算法利用有關單個字跡的數據來產生最終扭曲結果。

> 進一步對筆劃施加局部仿射變換並將它們疊加到合成圖像中，希望能夠學習出更適合新示例中常見變化的特徵。



Row 1: original images.

Row 2: global affine transforms.

Row 3: affine transforms on strokes.

Row 4: global affine transforms layered on top of stroke transforms.



Reference

<http://www.cs.toronto.edu/~gkoch/files/msc-thesis.pdf>

Thanks for listening