Advanced Institute for Artificial Intelligence

N-shot Learning

Advanced Institute for Artificial Intelligence – AI2

https://advancedinstitute.ai

Agenda

Agenda

- Introdução
- Aprendizagem tradicional x aprendizagem por diferenciação

N-shot learning

- As redes neurais convolucionais profundas tornaram-se os métodos mais avançados para tarefas de classificação de imagens.
- A limitaçào é que eles exigem muitos dados rotulados. Em muitos aplicativos, a coleta de tantos dados às vezes não é viável.
- N-Shot Learning é uma técnica para atacar tal limitação

N-shot learning

Classificação padrão

- imagem de entrada é alimentada em uma série de camadas e, finalmente, na saída, geramos uma distribuição de probabilidade em todas as classes (normalmente usando um Softmax)
- Por exemplo, se estamos tentando classificar uma imagem como gato ou cachorro ou cavalo ou elefante, para cada imagem de entrada, geramos 4 probabilidades, indicando a probabilidade da imagem pertencer a cada uma das 4 classes.

N-shot learning

Classificação padrão

- Durante o processo de treinamento, exigimos um grande número de imagens para cada classe (gatos, cães, cavalos e elefantes).
- Se a rede for treinada apenas nas 4 classes de imagens acima, não podemos esperar testá-la em nenhuma outra classe, como zebra por exemplo
- Para que o modelo classifique imagens de zebra, precisamos primeiro obter muitas imagens da zebra e, em seguida, devemos treinar novamente o modelo.

N-shot learning

Desafios para montar problemas com base no modelo tradicional de classificação

- Em alguns cenários obter um número de amostras significativas que permita uma classificação adequada não é trivial
  - Obter fotos de certas espécies de animais ou plantas
  - Obter resultados de avaliação de experimentos médicos
- Em outros casos as categorias são muito dinâmicas
  - Reconhecimento facial para funcionários de uma empresa
- N-shot Learning é uma técnica apropriada a problemas com classes que possuem poucos dados

N-shot learning

Exemplos de aplicações N-shot

- Reconhecimento facial de funcionários de uma empresa
  - Para treinar esse sistema, primeiro exigimos muitas imagens diferentes de cada uma das 10 pessoas na organização que podem não ser viáveis. (Imagine se você estiver fazendo isso para uma organização com milhares de funcionários).
  - E se uma nova pessoa ingressar ou sair da organização? Você precisa coletar dados novamente e treinar novamente o modelo inteiro
- Omniglot dataset: classificar um caracter de acordo ao alfabeto linguístico que pertence

N-shot learning

Agora vamos entender como abordamos esse problema usando n-shot learning

- No lugar de classificar diretamente uma imagem de entrada (teste), essa rede obtém uma imagem de referência como entrada e produzirá uma pontuação de similaridade indicando as chances de as duas imagens de entrada pertencerem ao grupo.
- Normalmente, a pontuação de similaridade é calculada entre 0 e 1 usando uma função sigmóide; em que 0 indica sem similaridade e 1 indica similaridade completa
- A rede está aprendendo uma função de similaridade, que recebe duas imagens como entrada e retorna o quanto são semelhantes.

N-shot learning

Como isso resolve os dois problemas que discutimos acima?

- Para treinar essa rede poucas instâncias são suficientes
- Uma vantagem é adequar-se dinamicamente a entrada, no exemplo de reconhecimento facial, caso uma nova face precise ser identificada (um novo funcionário na empresa), apenas uma imagem desse funcionário será necessária
- Usando essa imagem como referência, a rede calculará a semelhança para qualquer nova instância apresentada a ela. Assim, dizemos que a rede prevê a pontuação de uma só vez

N-shot learning

Vamos entender como podemos mapear esse problema em uma tarefa de aprendizado supervisionado em que nosso conjunto de dados contém pares de (Xi, Yi) em que 'Xi' é a entrada e 'Yi' é a saída.

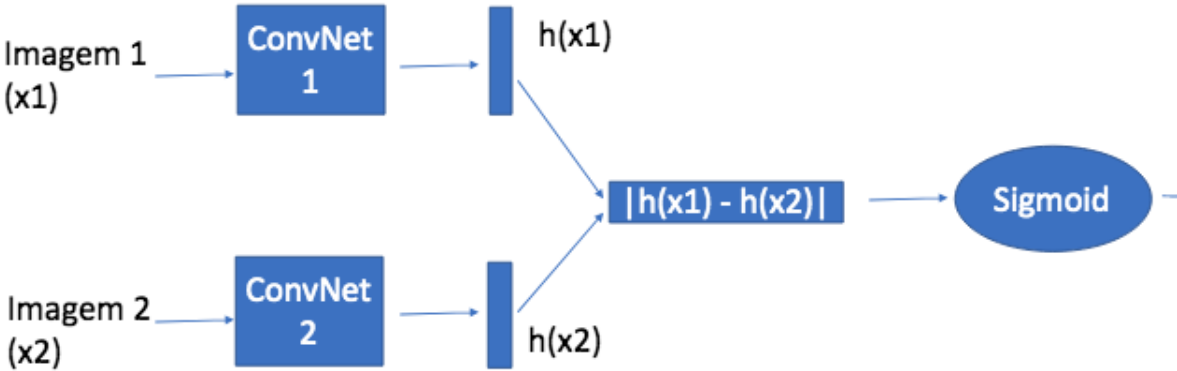
- A entrada para o nosso sistema será um par de imagens e a saída será uma pontuação de similaridade entre 0 e 1.
- Xi = Par de imagens
- Yi = 1; se ambas as imagens forem similares
- Yi = 0; se ambas as imagens forem diferentes entre si

N-shot learning

Redes siamesas

- Para treinar esse modelo é necessário gerar pares aleatórios de imagens e a variável alvo indicando se pertencem ou não ao mesmo grupo
- Os batches precisam ser criados com conjuntos de pares
- Os pares de imagens são submetidos a uma rede neural definda como siamêsa

AutoEncoder



N-shot learning

Redes Siamesas

- São duas redes identicas que levam a uma camada que as une
- Basicamente, eles compartilham os mesmos parâmetros. As duas imagens de entrada (x1 e x2) são passadas através de casa ConvNet para gerar um vetor de recurso de comprimento fixo para cada (h (x1) eh (x2)).
- Com base nesse modelo podemos formular a seguinte hipótese:
  - Se as duas imagens de entrada pertencerem ao mesmo caractere, seus vetores de características também deverão ser semelhantes
  - Se as duas imagens de entrada pertencerem a caracteres diferentes, suas vetores de recursos também serão diferentes.

N-shot learning

- A diferença absoluta em termos de elementos entre os dois vetores de características deve ser muito diferente nos dois casos acima.
- A pontuação de similaridade gerada pela camada sigmóide de saída também deve ser diferente nesses dois casos
- Essa é a idéia central por trás das redes siamesas.

N-shot learning

validação

- O modelo gerado ao fazer a predição, recebe duas imagens e retorna uma métrica de similaridade
- Isso não é suficiente para saber se o modelo é robusto o suficiente para diferenciar classes
- Para isso utilizamos conjuntos de dados que contém pares de imagens diferentes e um par do mesmo grupo

N-shot learning

A idéia do método de validação é a seguinte:

- definimos um tamanho N e criamos N pares com imagens diferentes entre si e apenas um desses pares com duas imagens iguais
- Se o modelo retorna 1 para imagem que faz parte do mesmo conjunto, dizemos que é uma predição correta, senáo é uma predição incorreta
- Depois de repetir esse processo K vezes, é possível calcular o nível de acurácia do modelo
- Quando maior o número N mais difícil do modelo alcancar acurácia alta