Zero shot learning

2017-01-04

Presented by Cheng-Hao Tu

Outline

- Motivation
- Introduction
- Related work
- Experiment
- Summary and future plan

Motivation

- The number of object categories can be huge in the open world
  - Pure supervised learning approach needs huge amount annotations
  - Training huge amount data needs huge amount of computation resources
- The situation becomes severe in fine-grained classification

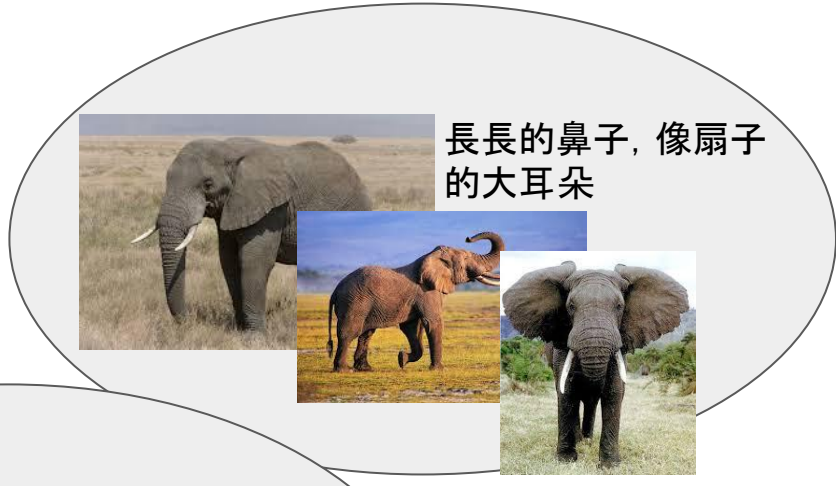


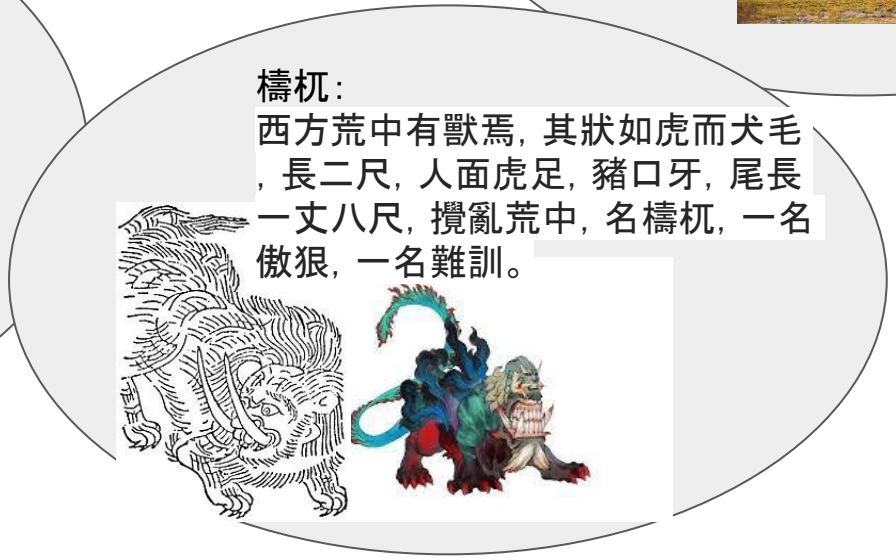

Introduction

- Zero-shot learning
  - To recognize classes that have no training samples
  - In training time:
    - Seen classes (train classes) and their descriptions
    - Seen classes training samples
  - In testing time:
    - Unseen classes and their descriptions
    - Given a new input sample, predict its class

Introduction (cont'd)

- In training time





Introduction (cont'd)

- In testing time

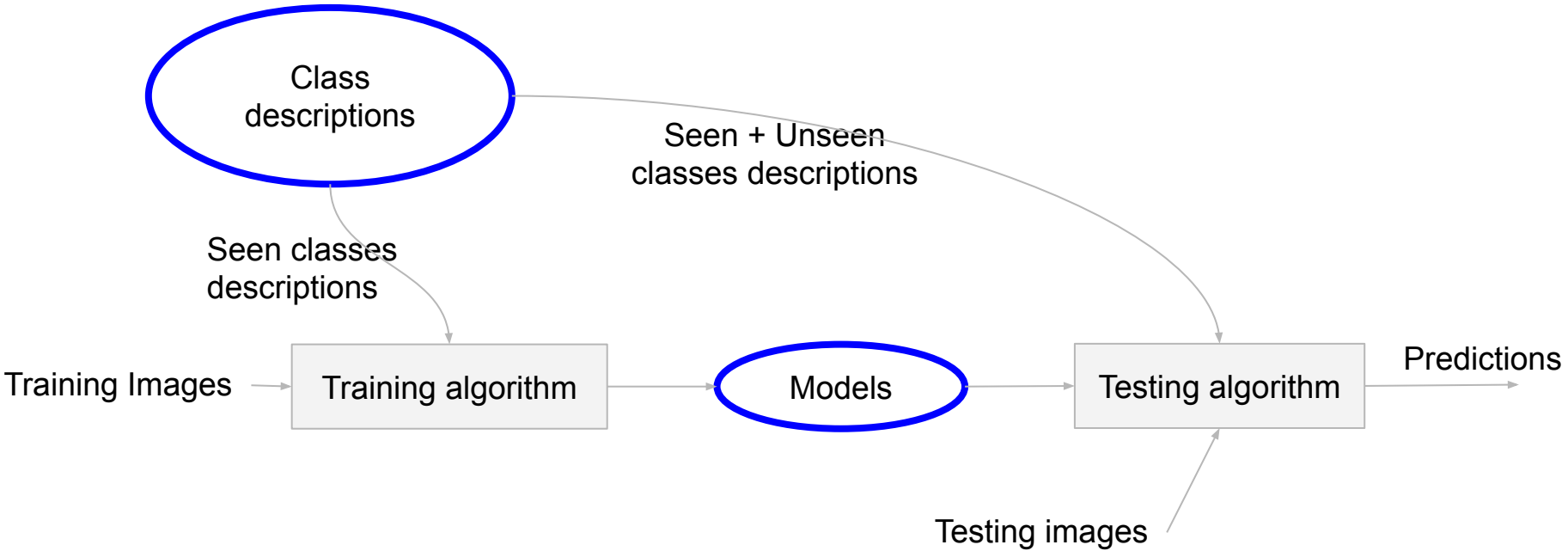
獅子的身體，人的臉

檮杌：
西方荒中有獸焉，其狀如虎而犬毛，長二尺，人面虎足，豬口牙，尾長一丈八尺，攪亂荒中，名檮杌，一名傲狠，一名難訓。

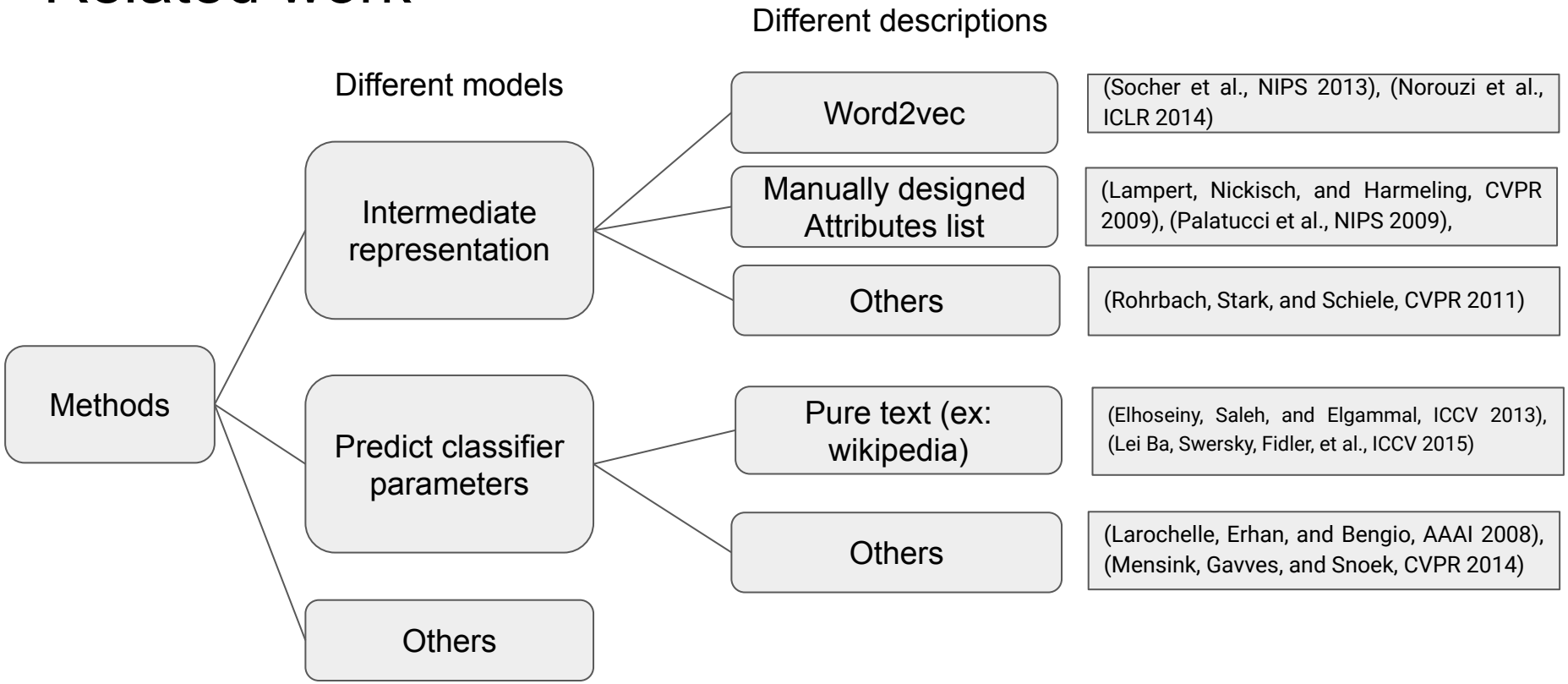
窮奇：
窮奇似牛而狸尾，尾長曳地，其聲似狗，狗頭人形，鉤爪鋸牙。

長長的鼻子，像扇子的大耳朵

Introduction (cont'd)

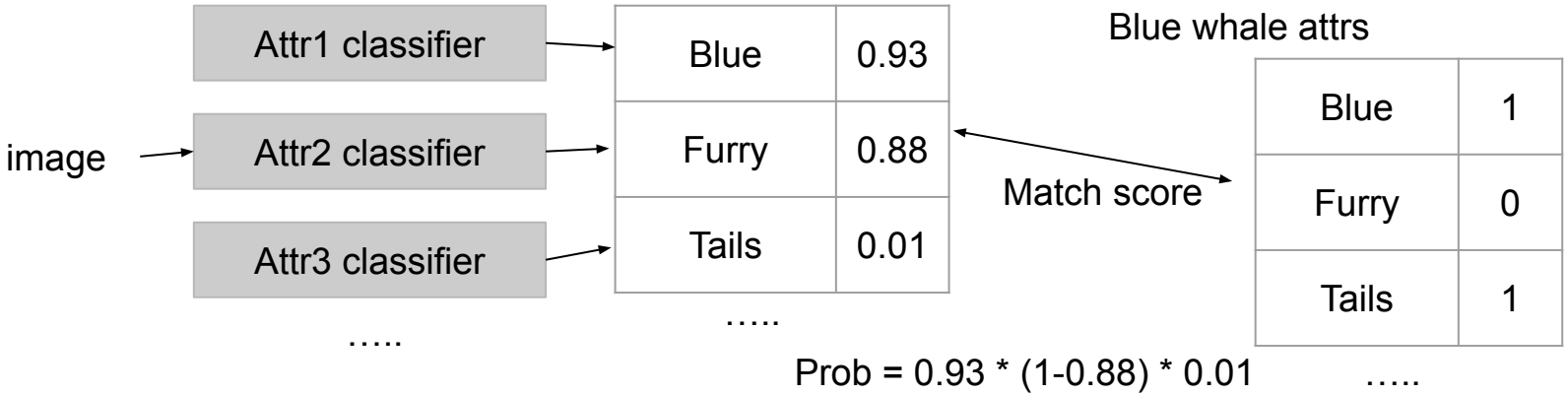


Related work



Related work (cont'd)

- Intermediate representation (IR)
  - Attributes list
  - Word2vec



Related work (cont'd)

- Predict classifier parameters
  - Pure text descriptions

Pure supervised learning case:

$$y_{i_1} = f_{w_1}(x_i),\ y_{i_2} = f_{w_2}(x_i), \ldots$$

These classifiers are trained using training data.
But what about $f_{w_u}(\cdot)$?

To determine whether sample x_i is class 1 $f_{w_1}(x_i)$

Zero-shot learning case:

$$y_{i_1} = f_{w_1}(x_i),\ y_{i_2} = f_{w_2}(x_i), \ldots \qquad f_{w_u}(\cdot) = f_{g_s(\ descriptions\ of\ unseen\ class\ )}$$

$$w_1 = g_s(\ descriptions\ of\ class1\ ) \qquad w_2 = g_s(\ descriptions\ of\ class2\ )$$

To determine whether sample x_i is class 1 $f_{g_s(\ descriptions\ of\ class1\ )}(x_i)$

Experiment

- Dataset: Animals with Attributes (AwA)
- 50 animals classes (separate to 40 seen classes, 10 unseen classes)
  - Seen classes: antelope, grizzly_bear, killer_whale, beaver, dalmatian, horse, german_shepherd, blue_whale, siamese_cat, skunk, mole, tiger, moose, spider_monkey, elephant, ……..
  - Unseen classes: chimpanzee, giant_panda, leopard, persian_cat, pig, hippopotamus, humpback_whale, raccoon, rat, seal
- No raw images (only provide extracted features) !!!
- Have 85 attributes
- No Pure text descriptions (I extract paragraph from wikipedia for each classes)

Experiment (cont'd)

- Two IR approaches modified from (Lampert, Nickisch, and Harmeling, CVPR 2009), (Socher et al., NIPS 2013)
  - Word2vec (google) and Attributes list (with 85 attributes provided by the dataset)
  - Using Neural network as model
- One "Predict classifier parameters" approaches (Lei Ba, Swersky, Fidler, et al., ICCV 2015)
  - Predict linear classifier weight on AwA only (since the dataset didn't provide feature map)
  - Using tf-idf as description features
  - Using vgg fc7 and image features

CUB200_2010 dataset		
	fc	fc+conv
mean top 1 acc	0.159 (0.17)	0.151
mean top 5 acc	0.334 (0.38)	0.330 (0.25)

AwA dataset Seen samples: 4859, Unseen samples: 6180, Total samples: 11039, 40 seen classes, 10 unseen classes

Top5 acc	Intermediate representation						Predict classifier parameters		
	Attributes			Google Word2Vec			Fc		
	Seen samples	Unseen samples	Total samples	Seen samples	Unseen samples	Total samples	Seen samples	Unseen samples	Total samples
Seen classes + Unseen classes	0.98	0.52	0.72	0.95	0.10	0.47	0.97	0.23	0.56
Unseen classes	-	0.91	-	-	0.47	-	-	0.79	-

AwA dataset Seen samples: 4859, Unseen samples: 6180, Total samples: 11039, 40 seen classes, 10 unseen classes

Top1 acc	Intermediate representation						Predict classifier parameters		
	Attributes			Google Word2Vec			Fc		
	Seen samples	Unseen samples	Total samples	Seen samples	Unseen samples	Total samples	Seen samples	Unseen samples	Total samples
Seen classes + Unseen classes	0.92	0.11	0.46	0.87	0.00	0.38	0.85	0.00	0.37
Unseen classes	-	0.62	-	-	0.17	-	-	0.20	-

Summary

- Briefly introduce the zero shot learning and some previous solutions
- Experiment on Animals with Attributes dataset using three methods
- Future plan:
  - Focus on face expression project