Expert Systems With Applications 42 (2015) 9279–9283

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Recognizing unknown objects with attributes relationship model

Wai Lam Hoo, Chee Seng Chan*

Centre of Image and Signal Processing, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Keywords:
Object recognition
Zero-shot learning
Attributes
Image hierarchy

ABSTRACT

Generally, training images are essential for a computer vision model to classify specific object class accurately. Unfortunately, there exist countless number of different object classes in real world, and it is almost impossible for a computer vision model to obtain a complete training images for each of the different object class. To overcome this problem, zero-shot learning algorithm was emerged to learn unknown object classes from a set of known object classes information. Among these methods, attributes and image hierarchy are the widely used methods. In this paper, we combine both the strength of attributes and image hierarchy by proposing Attributes Relationship Model (ARM) to perform zero-shot learning. We tested the efficiency of the proposed algorithm on Animals with Attributes (AwA) dataset and manage to achieve state-of-the-art accuracy (50.61%) compare to other recent methods.

1. Introduction

Object recognition is one of the active research areas in the computer vision community due to its usefulness in real-life applications, ranging from content-based image retrieval system such as search engines over the internet (Wu, Jin, & Jain, 2013), to video surveillance system to identify uncommon or suspicious objects in a selected area (Lee & Nevatia, 2014; Lim, Tang, & Chan, 2014). A well generalized object recognition system will greatly relax human efforts in identifying objects that have very minor difference in their appearance, but belonging to the same object category, for example by modelling a given object class by a set of modes deduced by a multi-finite mixture model (Bdiri, Bouguila, & Ziou, 2014; Bourouis, Mashrgy, & Bouguila, 2014).

However, when the numbers of distinct real-world objects grow increasingly large, it is very hard to have a computer vision model that is able to classify all of them. Besides, state-of-the-art object recognition algorithm always require a minimum number of samples from each object class to learn the difference between them. To add a new object class in the model after the learning process, the whole model will need to be retrained, and this is a tedious job. Therefore, zero-shot learning approach emerged (Frome et al., 2013; Lampert, Nickisch, & Harmeling, 2009, 2014; Palatucci, Pomerleau, Hinton, & Mitchell, 2009; Parikh & Grauman, 2011; Rohrbach, Stark, & Schiele, 2011), where it will be able to categorize unknown object classes from existing samples of other classes, utilizing the semantic relationship between the unknown object class and the existing object classes. For example, if three object classes need to be learned for the classification model, and one of the object class is unknown (due to no training samples), the other two known object classes in the recognition system will be utilized to find the characteristics of the unknown object class. Then, the classification model will be able to classify all three object classes. And retraining the full model is no longer needed.

Since then many research have focused on zero-shot learning tasks. Palatucci et al. (2009) were the first who initiated the zero-shot learning paradigm, to learn semantic output codes classifier that learns semantic properties of known classes to predict the unknown classes. Lampert et al. (2009, 2014) used Direct Attributes Prediction (DAP) model and Indirect Attributes Prediction (IAP) model, utilizing the attributes information. Whereas, Parikh and Grauman (2011) proposed relative attributes that further enhance zero-shot learning by introducing relative relationship between object classes using attributes, in contrast to binary attributes approach used in Lampert et al. (2009, 2014). There are other lines of research works (Frome et al., 2013; Hoo & Chan, 2013; Rohrbach et al., 2011) that favor image hierarchy approach finding the relationships between unknown object classes with existing known object classes. These approaches either rely on semantic information from WordNet, mining information from un-annotated data, or building specific Coarse Class-Fine Class lookup table. In more recent works, Fu, Hospedales, Xiang, and Gong (2014) proposed M2LATM that defines semi-latent attributes space, by using user-defined and latent attributes in one framework. Besides, Fu, Hospedales, Xiang, and Gong (2015a) used transductive multi-view embedding and heterogeneous multi-view label propagation method to overcome the known problems in zero-shot learning namely the projection domain shift and prototype sparsity.

* Corresponding author. Tel.: +0060 379676433.
E-mail addresses: wailam88@siswa.um.edu.my (W.L. Hoo), cs.chan@um.edu.my (C.S. Chan).

http://dx.doi.org/10.1016/j.eswa.2015.07.049
0957-4174/

In addition, Liu, Zhang, and Chen (2014) proposed to learn attributes relation and attributes classifier jointly in a common objective function, while Fu, Xiang, Kodirov, and Gong (2015b) suggested to use semantic manifold distance to project semantic embedding space and recognize unknown object classes. All these aforementioned methods only use either attributes or image hierarchy.

In this paper, we proposed to combine the benefits of both attributes and image hierarchy. Specifically, we build an Attributes Relationship Model (ARM) to perform zero-shot learning, based on the hierarchical class concept in Hoo and Chan (2013, 2015). Our contribution is, instead of using the Coarse Class - Fine Class relation as in their paper, we proposed to use attributes to build the relationship model. Our intuition is given the attributes of each unknown object class, the known object classes that have similar attributes with the unknown object class will have a stronger relationship. Since each attributes only represent characteristics in part of the image, we group the known object classes with high correlation to the unknown object classes based on their attributes. In short, we have a centralized relationship model that infer which known object classes is most correlated to the specific unknown object class. This is different from the relationship used in DAP and IAP (Lampert et al., 2009, 2014), which is not class-specific. These advantages enable the proposed method to enhance the zero-shot learning performance, where we achieve state-of-the-art results (50.41%) in Animal with Attributes (AwA) dataset.

This paper is organized as follows: we first formulate the proposed ARM model in Section 2. After that, we compare our relationship model with the current state-of-the-arts in Section 3. We then discuss our findings in Section 4, and conclude the paper in Section 5.

2. Attributes relationship model (ARM)

The proposed ARM aims to solve the zero-shot learning problem. Conventional learning models need to have at least one image sample of each object classes to learn their model. However, zero-shot learning allows missing training images on selected object class(es), denoted as the unknown object class. Attributes, in here, helps to relate unknown object class with the known object classes, because attributes are shared among all object classes (as in Lampert et al. (2009, 2014); Parikh and Grauman (2011).

2.1. Attributes

Attributes can be interpreted as visual or conceptual elements that exist in different objects. Taking animals as example, Animals with Attributes (AwA) dataset (Lampert et al., 2009) relates 50 object classes with 85 attributes, either in binary or in real-value to describe their association strength, as shown in Fig. 1. in class-attribute matrix (that explains the relationship between object classes and attributes). Brighter color in the matrix indicates the class-attributes pair has stronger association, while darker color means weaker association. These attributes could be based on visual appearance (e.g. black, white, furry, paws etc), behavior description (e.g. strong, fast, active etc) and habitat (e.g. forest, bush, desert etc). Lampert et al. (2009, 2014) used these attributes to investigate unknown classes by proposing DAP and IAP. Based on their findings, attributes proves to be a better intermediate feature representation compared to other low-level representation such as edge, because attributes describe the object characteristics in a semantic manner. However, a significant amount of computational time is needed to learn the DAP or IAP.

2.2. ARM learning

The proposed ARM builds a simplified version of relationship (compared to DAP and IAP (Lampert et al., 2009, 2014)) that is class-specific, where it will describe a given unknown object class, depending on the information gathered (e.g. attributes) from the known object classes. Formally, attributes are denoted as $A = \{a_1, a_2, \ldots, a_n\}$. Each object class c will associate with a set of $a \subset A$. Given an unknown object class c_u, we first identify attributes that belonged to the object class $a_i \in c_u$. Then, we increase the correlation score for the known object classes c_s that shares $a \subset c_s$. After that we find other c_s that consist of $a_n \in c_s$ and increase the correlation score by either using the binary value or real-valued association strength:

$$\mathrm{score}(c_s|c_u) = \frac{1}{r}\sum_{i=1}^{r}\begin{cases} a_i a_n; & \text{if } a_i \in c_u, a_n \in c_s \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where a_i and a_n indicate the scores of the attributes, as described in the class-attribute matrix, and r is the number of attributes that exist in c_u. Intuitively, when the binary attributes are used, a_i and a_n will be either 1 or 0. When real-valued attributes are used, a_i and a_n act as a weight that describe how important is the attribute to c_u and c_s respectively, so that c_s possesses the same attributes will have higher relative correlation with the c_u based on a_i and a_n. We show the generated ARM in Tables 1 and 2, based on the binary attributes and real-valued attributes given in Fig. 1. There are a total of 10 c_u

Table 1
Attributes Relationship Model (ARM) that is generated using binary attributes as in Fig. 1(a).

c_u	Strongly related c_s (First 5)
Chimpanzee	Rhinoceros ≻ Elephant ≻ Deer ≻ Blue Whale ≻ Moose
Giant Panda	Squirrel ≻ German Shepherd ≻ Gorilla ≻ Cow ≻ Giraffe
Leopard	Moose ≻ Elephant ≻ Deer ≻ Rhinoceros ≻ Giraffe
Persian Cat	Wolf ≻ Otter ≻ Tiger ≻ Lion ≻ German Shepherd
Pig	Grizzly Bear ≻ Deer ≻ Rhinoceros ≻ Blue Whale ≻ Skunk
Hippopotamus	Skunk ≻ Dalmatian ≻ Blue Whale ≻ Deer ≻ Elephant
Humpback Whale	Fox ≻ Siamese Cat ≻ Bat ≻ Killer Whale ≻ Beaver
Raccoon	Lion ≻ Rhinoceros ≻ Deer ≻ Skunk ≻ Elephant
Rat	Squirrel ≻ German Shepherd ≻ Gorilla ≻ Zebra ≻ Dolphin
Seal	Bat ≻ Fox ≻ Siamese Cat ≻ Killer Whale ≻ German Shepherd

Table 2
Attributes Relationship Model (ARM) that is generated using real-valued attributes as in Fig. 1(b).

c_u	Strongly related c_s (First 5)
Chimpanzee	Moose ≻ Elephant ≻ Deer ≻ Phinoceros ≻ Blue Whale
Giant Panda	Squirrel ≻ German Shepherd ≻ Gorilla ≻ Cow ≻ Sheep
Leopard	Moose ≻ Elephant ≻ Deer ≻ Blue Whale ≻ Grizzly Bear
Persian Cat	Lion ≻ Otter ≻ Wolf ≻ Mole ≻ Chihuahua
Pig	Moose ≻ Grizzly Bear ≻ Deer ≻ Elephant ≻ Blue Whale
Hippopotamus	Blue Whale ≻ Moose ≻ Skunk ≻ Elephant ≻ Deer
Humpback Whale	Fox ≻ Killer Whale ≻ Bat ≻ Siamese Cat ≻ Polar Bear
Raccoon	Lion ≻ Moose ≻ Deer ≻ Mole ≻ Rhinoceros
Rat	Squirrel ≻ German Shepherd ≻ Gorilla ≻ Walrus ≻ Cow
Seal	Fox ≻ Killer Whale ≻ Bat ≻ Siamese Cat ≻ Polar Bear