Prior Knowledge about Attributes: Learning a More Effective Potential Space for Zero-Shot Recognition

Chunlai Chai

Zhejiang Gongshang University Hangzhou, China ccl@mail.zjgsu.edu.cn

Yukuan Lou

Zhejiang Gongshang University Hangzhou, China louyukuan@gmail.com

Shijin Zhang

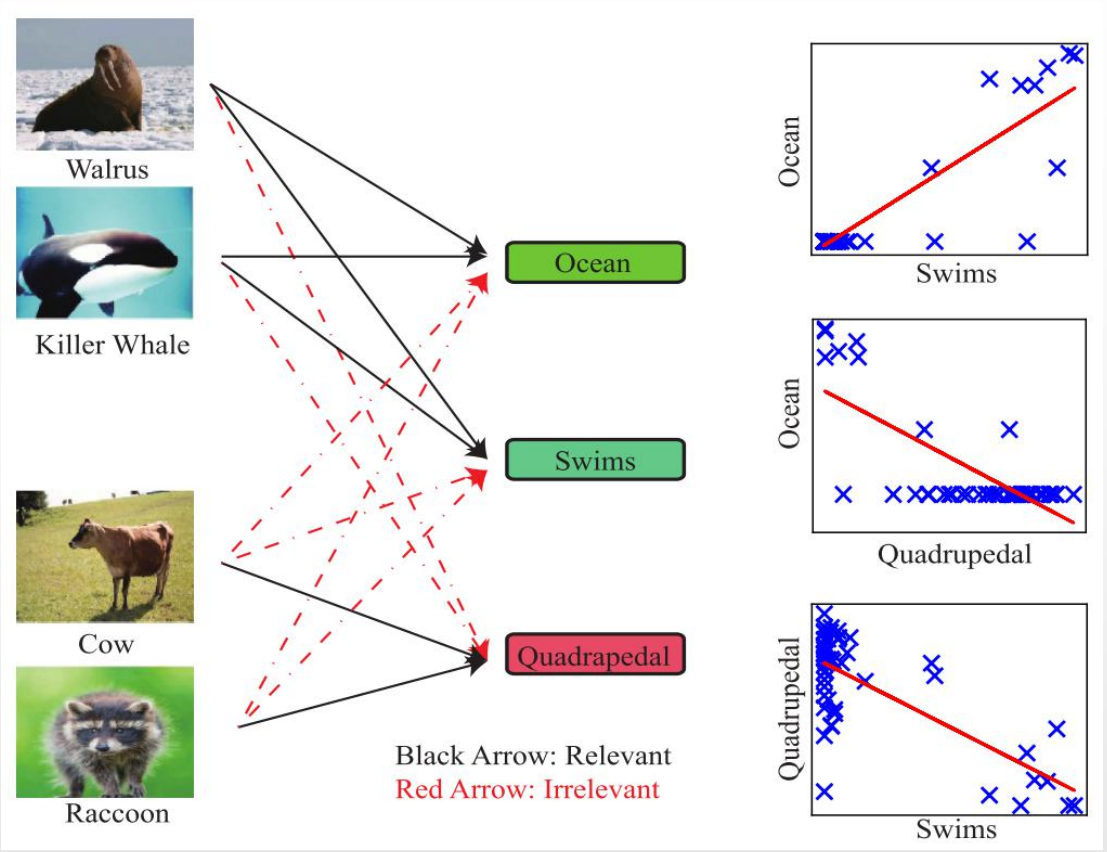
Zhejiang Gongshang University Hangzhou, China zhangshijin0304@gmail.com

Ming Hua

Oakland University State of Michigan, US ming@oakland.edu

1. INTRODUCTION

Zero-shot learning (ZSL) aims to recognize unseen classes accurately by learning seen classes and known attributes, but correlations in attributes were ignored by previous study which lead to classification results confused. To solve this problem, we build an Attribute Correlation Potential Space Generation (ACPSG) model which uses a graph convolution network and attribute correlation to generate a more discriminating potential space. Combining potential discrimination space and user-defined attribute space, we can better classify unseen classes. Our approach outperforms some existing state-of-the-art methods on several benchmark datasets, whether it is conventional ZSL or generalized ZSL.



An illustrative diagram of semantic attributes correlation. It can be seen from the figure that if an animal has a certain attribute, it is likely to also have an attribute related to it; otherwise, it is likely not to have an attribute not related to it.

2. RELATED WORK

- A. Zero-Shot Learning

According to previous research, there are three types of models in ZSL: (1) Class-Inductive Instance-Inductive setting, it means training the model using only the trainable instances and the set of seen labels, (2) Class-Transductive InstanceInductive setting, This means to train models using trainable instances and a set of seen labels, as well as a set of unseen labels. (3) Class-Transductive Instance-Inductive setting, it means to train the model using trainable instances and seen label sets, as well as unseen label sets and corresponding unlabeled test sets.

- B. Graph Neural Networks

Graph convolution was first proposed to extend CNN to graphs and to directly process graph-type data. CNN generally acts on Euclidean space, and cannot directly act on nonEuclidean space. Many important data sets are stored in the form of graphs in reality, such as social network information, knowledge map, protein network, the World Wide Web and so on. The form of these graph networks is not like an image. It is composed of a neatly arranged matrix but is unstructured information. CNN cannot be used for feature extraction, but graph convolution can be applied here.

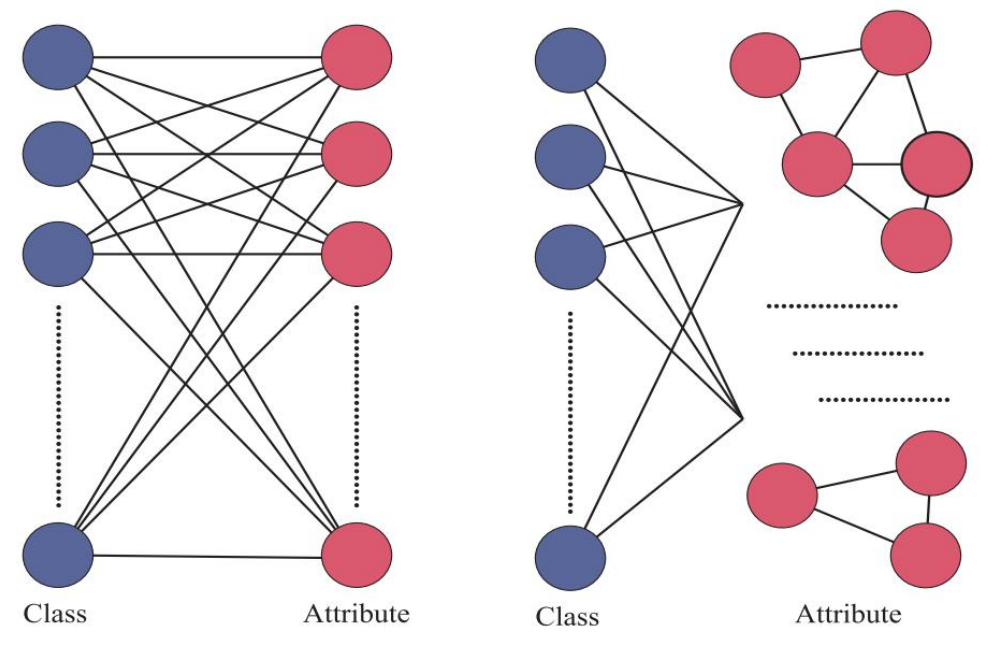
The core of graph convolution is that each node in the graph is affected by neighbor nodes and further points at any time, so it constantly changes its state until the final balance. The nodes closer to the target node have a greater influence on the target node. GCN has subtly designed a method for extracting features from graph data so that we can use these features to perform node classification, graph classification, and edge prediction on graph data. It is versatile that we can get embedded representations of graphs in this way.

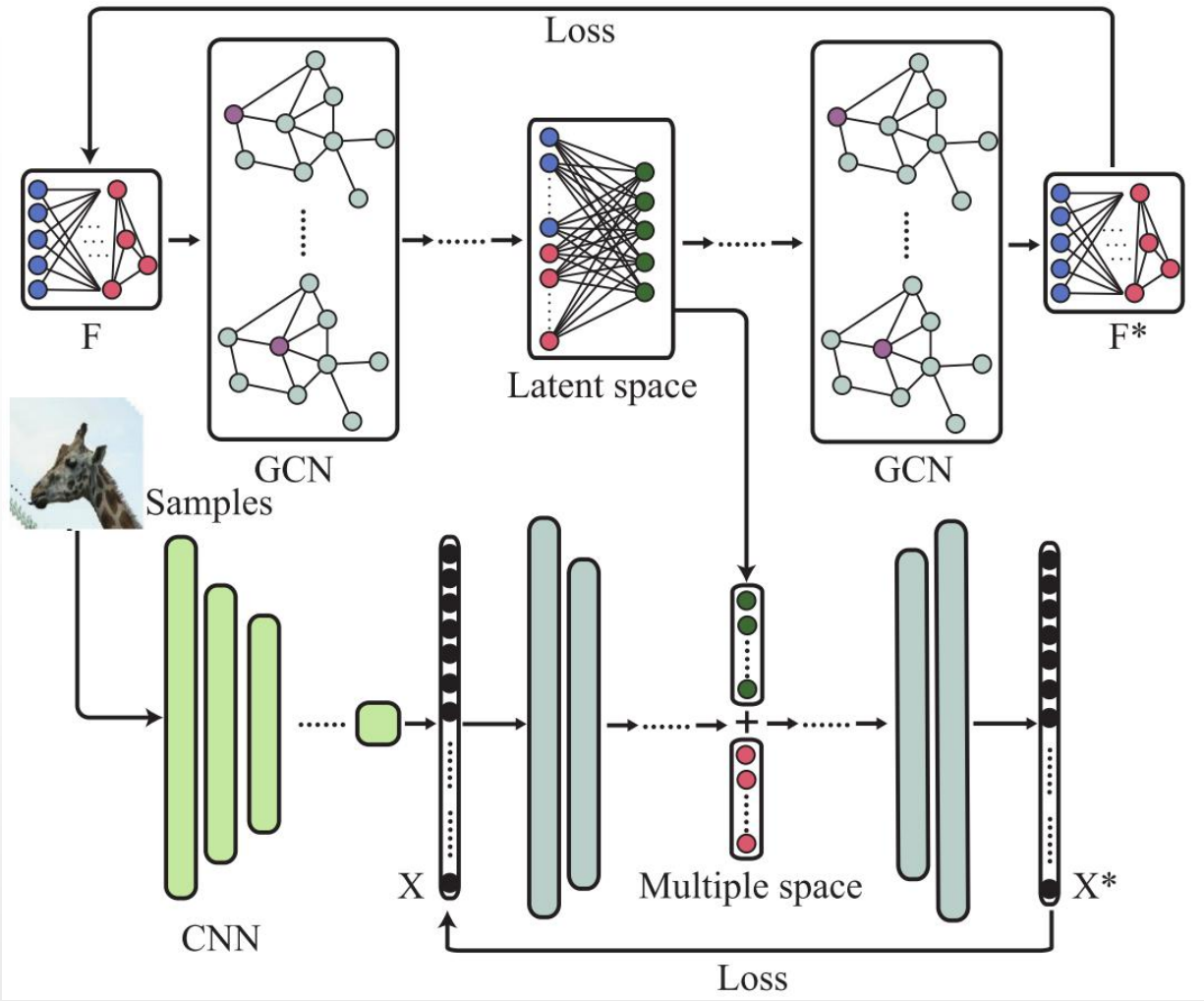
- C. AutoEncoder

An autoencoder is an unsupervised neural network model. It can learn the hidden features of the input data, which can be called encoding. At the same time, the new input features can be used to reconstruct the original input data, which is called decoding. Intuitively, auto-encoders can be used to reduce the feature dimension is similar to principal component analysis, but its performance is stronger than PCA. This is because neural network models can extract more efficient new features. In addition to feature dimensionality reduction, new features learned by the autoencoder can be input into a supervised learning model, so the autoencoder can be used as a feature extractor. As an unsupervised learning model, autoencoders can also be used to generate new data that is different from the training samples, such as variational autoencoders.

3. OUR MODEL

- We solve this problem by graph convolutional networks(GCN) , GCN uses the correlation between class nodes and semantic attribute nodes to generate a latent space to help identify unseen classes. Before this research, the bipartite graph has been used to represent the correlation between ZSL nodes, as it is shown in right figure, but the bipartite graph ignores the correlation in semantic attribute nodes. We propose a new graph model to replace the bipartite graph, covering the correlation between semantic attribute nodes, thus generating better potential space. A new ZSL framework called Attribute Correlation Potential Space Generation (ACPSG) model consists of two parts, the first part generates latent discrimination attribute space from GCN, the second part maps the visual features of the unseen classes into userdefined attribute space and latent discrimination attribute space through an autoencoder . In the end, combining multiple spaces, we can consider both the UA and LA spaces to perform ZSL prediction.





Overall illustration of the framework proposed in this paper. At the first stage, we added the correlation between attributes as a prior knowledge, using a graph convolution model to generate a latent discernment space. In the second stage, we use autoencoders to map visual features into multiple spaces and learn a reliable decoder.

4. EXPERIMENTS

- Conventional Zero-Shot Learning

we evaluated our model in detail on three benchmark datasets (AwA2, CUB, aPY). Experimental results show that model performs well and outperforms some advanced models in some results.

THE EXPERIMENTAL RESULTS ON THE CONVENTIONAL ZSL. HERE THE PS AND THE SS SEPARATELY REFER TO THE PROPOSED SPLIT AND THE STANDARD SPLIT. THE BEST RESULT IS MARKED IN BOLD FONT. NONE MEANS NO POTENTIAL DISCERNMENT SPACE IS USED. $S->V$ MEANS USE EQ. 18 TO CALCULATE ACCURACY, $V->S$ MEANS USE EQ. 17 TO CALCULATE ACCURACY.

Method	AwA2		CUB		aPY	
	SS	PS	SS	PS	SS	PS
DAP [41]	58.7	46.1	37.5	40.0	35.2	33.8
CONSE [42]	67.9	44.5	36.7	34.3	25.9	26.9
ALE [6]	80.3	62.5	53.2	54.9	-	39.7
ESZSL [10]	75.6	58.6	55.1	53.9	34.4	38.3
SJE [7]	69.5	61.9	55.3	53.9	32.0	32.9
SYNC [43]	71.2	46.6	54.1	55.6	39.7	23.9
SAE [23]	80.2	54.1	33.4	33.3	**55.4**	8.3
SE-ZSL [16]	80.8	69.2	60.3	59.6	-	-
ZSKL [28]	-	**70.2**	-	57.1	-	**45.3**
F-CLSWGAN [17]	-	-	-	61.5	-	-
DCN [44]	-	-	55.6	56.2	-	43.6
PSRZSL [45]	-	63.8	-	56.0	-	38.4
None (S->V)	78.1	57.9	59.2	58.6	41.7	23.4
ACPSG (S->V)	79.5	66.1	61.3	59.1	45.1	27.3
None (V->S)	79.8	48.9	58.1	57.4	41.3	24.4
ACPSG (V->S)	**82.8**	49.8	**62.7**	**60.4**	47.2	27.5

- Generalized Zero-Shot Learning

we tested our model on three benchmark datasets, it shows a good performance on AwA2 and CUB, But not so good on aPY, we guessed that it was caused by the insufficient fine-grained of the learned distribution. But in general, our model has good generalization capability.

THE EXPERIMENTAL RESULTS ON THE GENERALIZED ZSL. S REPRESENTS THE ACCURACY OF SEEN CLASSES. U REPRESENTS THE ACCURACY OF UNSEEN CLASSES. H IS THE HARMONIC MEAN. THE BEST RESULT IS MARKED IN BOLD FONT.

Method	AwA2			CUB			aPY		
	S	U	H	S	U	H	S	U	H
CONSE [42]	**90.6**	0.5	1.0	**72.2**	1.6	3.1	**91.2**	0.0	0.0
CMT [46]	90.0	0.5	1.0	49.8	7.2	12.6	74.2	10.9	19.0
SJE [7]	73.9	8.0	14.4	59.2	23.5	33.6	55.7	3.7	6.9
ESZSL [10]	77.8	5.9	11.0	63.8	12.6	21.0	70.1	2.4	4.6
SYNC [43]	90.5	10.0	18.0	70.9	11.5	19.8	66.3	7.4	13.3
SAE [23]	82.2	1.1	2.2	54.0	7.8	13.6	80.9	0.4	0.9
LATEM [47]	77.3	11.5	20.0	57.3	15.2	24.0	73.0	0.1	0.2
ALE [6]	81.8	14.0	23.9	62.8	23.7	34.4	73.7	4.6	8.7
ZSKL [28]	82.7	18.9	30.8	52.8	21.6	30.6	76.2	10.5	18.5
PSRZSL [45]	73.8	20.7	32.3	54.3	24.6	33.9	51.4	13.5	21.4
DCN [44]	-	-	-	37.0	**25.5**	30.2	75.0	**14.2**	**23.9**
Ours	82.5	**23.1**	**36.1**	71.3	25.0	**37.0**	76.3	8.8	15.8

5. CONCLUSION

- In this paper, we put forward the concept of attribute correlation in ZSL, and explore the correlation in attribute nodes, it makes attribute nodes are interrelated rather than isolated. To use attribute correlation as a prior knowledge of ZSL, we propose the ACPSG model to make full use of the correlation between nodes. Specifically, our model learns multiple spaces that are more discernible than the original space. Using this method, we integrated attribute correlation into the ZSL model successfully. Besides, we have done a lot of experiments to verify the effectiveness of our model.
- In essence, the graph-based approach aims to model the interaction in entities. In our model, classes and attributes are regarded as different nodes in the graph, and edges are used to describe the correlation between the nodes so that the structural information between the various nodes is fully utilized. From all information we discussed, we use the graph convolutional networks to generate a more effective space for potential discrimination.
- In reality, we combine the latent discriminating space and the user-defined space into multiple spaces. We train the samples so that the visual features of the samples are mapped into multiple spaces, and the same class is clustered together and distributed reasonably.
- There are still many challenges in zero-sample learning. In the future, we will continue to develop ZSL models that based on graphs and attributes to give model better performance and generalization.