Learning Attribute Relation in Attribute-Based Zero-Shot Classification

Mingxia Liu[1,2], Songcan Chen[1], and Daoqiang Zhang[1]

[1] School of Computer Science and Engineering,
Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, P.R. China
[2] School of Information Science and Technology,
Taishan University, Taian, 271021, P.R. China
{mingxialiu,s.chen,dqzhang}@nuaa.edu.cn

Abstract. Recently, zero-shot learning has attracted increasing attention in computer vision community. One way of realizing zero-shot learning is by resorting to knowledge about attributes and object categories. Most existing attribute-centric approaches focus on attribute-class relation artificially derived by linguistic knowledge base or mutual information. In this paper, we aim to learn the attribute-attribute relation automatically and explicitly. Specifically, we propose to incorporate the attribute relation learning into attribute classifier design in a unified framework. Furthermore, we develop a new scheme for attribute-based zero-shot object classification, such that the learned attribute relation can be reused to boost the traditional attribute classifiers. Extensive experimental results demonstrate that our proposed method can enhance the performance of attribute prediction and zero-shot learning.

Keywords: attribute relation, attribute-based classification, zero-shot learning.

1 Introduction

In recent years, the problem of object classification in zero-shot scenarios (i.e. no training samples are available for the target classes) has attracted increasing attention in computer vision community [1-6]. This problem is challenging and on the other hand very useful, especially for such a real-world setting where the number of object classes is large but samples are unable or costly to be acquired. In contrast, high-level semantic attributes for each class can be obtained more conveniently, such as color and texture for arbitrary objects.

As the training and the test classes are commonly disjoint in zero-shot learning, researchers have developed several alternative methods for knowledge transfer from training to unseen test classes. Among them, attribute-centric methods based on high-level attributes have been proven effective in leveraging knowledge between attribute and object category experimentally [1-3].Existing attribute-centric methods focus on exploitation of the semantic relation between *attribute* and *object class* [4-6]. Likewise, intuitively, the relation between attributes of an object can also convey supplementary information from another aspect. However, in traditional methods,

J. Yang, F. Fang, and C. Sun (Eds.): IScIDE 2012, LNCS 7751, pp. 514–521, 2013.

attribute classifiers are trained separately with respect to individual attributes, and seldom consider the *attribute-attribute relation*.

In fact, there are often some correlation relations between attributes. Thus taking such relation into account in the process of training attribute classifiers should be helpful. For example, researchers in [2,4,13] have considered the attribute relation to some degree. However, the relations they adopted are largely derived artificially or predefined. In contrast, in this paper, we propose to train attribute classifiers and learn the attribute-attribute relation simultaneously in a unified objective function.

In this paper, we first propose a joint learning method for attribute classifiers coupled with the attribute-attribute relation. And then we develop a new attribute-based zero-shot learning scheme by incorporating the attribute relation learned in previous step, and demonstrate that it enhance the performance of attribute prediction.

The rest of this paper is organized as follows. Section 2 introduces related works of zero-shot and attributes relation learning. The proposed attribute relation learning (ARL) method is described in Section 3. In Section 4, we propose our attribute-based zero-shot object classification scheme with attribute relation incorporated. Extensive experiments are carried out in Section 5. Finally, conclusion is given in Section 6.

2 Related Works

To the best of our knowledge, the concept of zero-shot learning can at least be traced back to one of the early works proposed by Larochelle et al.[7]. It attempts to solve the problem of predicting novel samples that were unseen in the training data set. In [8, 9], researchers proposed methods to obtain the intermediate class description such as semantic knowledge base to perform zero-shot learning. In computer vision community, Farhadi et al.[3] described objects by their semantic or discriminative attributes, and revealed the potential to predict novel classes in zero-shot scenarios. In order to tackle the problem of learning with disjoint training and testing classes, Lampert et al.[14] proposed attribute-based classification, as well as Direct Attitude Prediction (DAP) and Indirect Attitude Prediction (IAP) to perform zero-shot learning. More recent related works can be found in [4, 5, 11, 12].

On the other hand, as for attribute-centric object classification, some researchers have demonstrated considering attribute-class relation can result in improved generalization performance [4-6, 13]. However, the attribute relation needs to be pre-computed, which are then used in a latent discriminative model for classification. Rohrbach et al.[2] used external linguistic knowledge bases and proper semantic relatedness to capture attribute-class relation, which requires extra expert knowledge for natural language processing. Siddiquie et al.[5] demonstrated modeling pairwise correlations between attributes brings better results in image ranking and retrieval.

Different from the previous works, we model the attribute relation in a totally new perspective. First, we learn an attribute covariance matrix that models the relation between attributes in the form of matrix variant normal distribution, motivated by multi-task relation learning [14-16]. Second, the attribute relation we learned can be incorporated into traditional classifiers separately and has been proven effective to enhance the performance of zero-shot learning in our experiments.

3 Proposed Approach

Suppose we are given a mid-level representation in the form of inventories of attributes $\{A_m\}_{m=1}^{M}$ for N object classes. Given a set of training images$\{x_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^d$ and class labels $l_i^m = \{1, \cdots, \mathrm{N}\}$ as well as attributes labels $y_i^m \in \{1,0\}$. Binary attributes are considered in this paper. For each attribute classifier, we learn a linear function $f_m(\mathbf{x}) = w_m^T \mathbf{x} + b_m$, where $\mathbf{W}$ is the weight vector and $\mathbf{b}$ is the bias term.

3.1 Problem Formulation

Since data for each attribute classifier is in the same pool, we denote as $\mathbf{X} = (x_1^1, \ldots, x_N^1, \ldots, x_1^M, \ldots, x_N^M)^T$, the attribute labels $\mathrm{y} = (y_1^1, \ldots, y_N^1, \ldots, y_1^M, \ldots, y_N^M)^T$ and the bias term $\mathbf{b} = (b_1, \ldots, b_M)^T$. Then the posterior distribution for $\mathbf{W}$ can be obtained through the prior and the likelihood in the following function [22]:

Let $N(\mathbf{m}, \mathbf{\Sigma})$ denotes multivariate normal distribution with mean$\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$. Given x_i^m, w_m, b_m and ε_m, the likelihood of y_i^m is given in the following form:

$$y_i^m \mid x_i^m, w_m, b_m, \varepsilon_m \sim N(\, w_m^T\, x_i^m + b_m,\ \varepsilon_m^2) \qquad (1)$$

Denote $\mathbf{I}_d$ as the $d \times d$ identity matrix, and the prior on $\mathbf{W} = (w_1, \ldots, w_\mathrm{M})$ is defined as

$$\mathbf{W} | \epsilon_m \sim \left(\prod_{m=1}^{\mathrm{M}} \mathrm{N}(w_m | \mathbf{0}_d,\ \epsilon_m^2 \mathbf{I}_d)\right) q(\mathbf{W}) \qquad (2)$$

where the first term is employed to control the column complexity and second one is for structure modeling of $\mathbf{W}$. As $\mathbf{W}$ is a matrix variable, the matrix variant normal distribution [17] is used as $q(\boldsymbol{W})$. So $q(\mathbf{W})$ can be defined as

$$q(\mathbf{W}) = \mathrm{MN}_{d \times m}(\mathbf{W} | \mathbf{0}_{d \times m}, \mathbf{I}_d \otimes \mathbf{\Omega}) \qquad (3)$$

Here, $\mathbf{I}_d$is a row covariance matrix modeling the features relation and $\mathbf{\Omega}$ is a column covariance matrix for the relation between w_m's.

Since data for each attribute classifier is in the same pool, we denote as $\mathbf{X} = (x_1^1, \ldots, x_N^1, \ldots, x_1^M, \ldots, x_N^M)^T$, the attribute labels $\mathrm{y} = (y_1^1, \ldots, y_N^1, \ldots, y_1^M, \ldots, y_N^M)^T$ and the bias term $\mathbf{b} = (b_1, \ldots, b_M)^T$. Then the posterior distribution for $\mathbf{W}$ can be obtained through the prior and the likelihood in the following function [22]:

$$p(\mathbf{W} | \mathbf{X}, \mathbf{y}, \mathbf{b}, \varepsilon, \epsilon, \mathbf{\Omega}) \propto p(\mathbf{y} | \mathbf{X}, \mathbf{W}, \mathbf{b}, \varepsilon) p(\mathbf{W} | \epsilon, \mathbf{\Omega}) \qquad (4)$$

By taking the negative logarithm of Eq. (4) and combing it with Eqs. (1)-(3), the maximum likelihood estimation of $\mathbf{\Omega}$ and $\mathbf{b}$, as well as the maximum a posterior estimation of W, can be obtained through the following: .

$$\min_{\mathbf{W}, \mathbf{b}, \mathbf{\Omega}} \sum_{m=1}^{\mathrm{M}} \frac{1}{\varepsilon_m^2} \sum_{i=1}^{\mathrm{N}} \left(y_i^m - (w_m^\mathrm{T} x_i^m + b_m) \right)^2 + \sum_{m=1}^{\mathrm{M}} \frac{1}{\epsilon_m^2} w_m^\mathrm{T} w_m$$
$$+ \mathrm{tr}(\mathbf{W} \mathbf{\Omega}^{-1} \mathbf{W}^\mathrm{T}) + d \ln |\mathbf{\Omega}| \qquad (5)$$

For convenience to optimize the problem Eq. (5), the last term $d\ln|\mathbf{\Omega}|$ can be replaced by the constraint $tr(\mathbf{\Omega}) = 1$ which is convex, with the same aim to restrict the complexity of $\mathbf{\Omega}$ .Then the model can be rewritten as the following form:

$$\min_{\mathbf{W},\mathbf{b},\mathbf{\Omega}} \sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N}\left(y_i^m - (w_m^T x_i^m + b_m)\right)^2 + \frac{\lambda_1}{2}tr(\mathbf{W}\mathbf{W}^T) + \frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^T)$$
$$\text{s.t. } \mathbf{\Omega} \geq 0,\ \ tr(\mathbf{\Omega}) = 1 \qquad (6)$$

where $\lambda_1 = \frac{2\varepsilon^2}{\epsilon^2}$, and $\lambda_2 = 2\varepsilon^2$. The constraint $\mathbf{\Omega} \geq 0$ in Eq. (6) is used to restrict $\mathbf{\Omega}$ as positive semi-definite because it is the attribute covariance matrix. The first term in Eq. (6) gives the empirical loss on the training data. And the second one is to penalize the complexity of $\mathbf{W}$. It is worth noting that the last term $tr(\mathbf{W}\mathbf{\Omega}^{-1}\mathbf{W}^T)$ is to model the relation between all attributes.

3.2 Alternating Optimization Algorithm

In Eq. (6), three variables to be optimized are jointly convex, which can be achieved through an alternating optimization method. The first step is to optimize $\mathbf{W}$ and $\mathbf{b}$ given a fixed $\mathbf{\Omega}$, and the second one is to optimize $\mathbf{\Omega}$ when $\mathbf{W}$ and $\mathbf{b}$ are fixed.

Optimizing W and b when Ω is fixed

As is shown in [16], the dual problem defined in Eq. (6) can be written as

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\boldsymbol{\alpha}^T\widetilde{\mathbf{K}}\boldsymbol{\alpha} - \sum_{m=1}^{M}\sum_{i=1}^{N}\alpha_i^m y_i^m \ \text{ s.t. } \ \sum_{i=1}^{N}\alpha_i^m = 0,\ \forall m,\ m = 1,\dots,M \quad (7)$$

where $\widetilde{\mathbf{K}} = \mathbf{K} + \frac{1}{2}\mathbf{\Lambda}$, $\boldsymbol{\alpha} = (\alpha_1^1, \dots \alpha_N^1, \dots, \alpha_1^M, \dots \alpha_N^M)^T$ and. Note that $\mathbf{K}$ is the kernel matrix on all data points for all attributes classifiers, whose element is $k(x_{i1}^{m1}, x_{i2}^{m2}) = \mathbf{e}_{m1}^T\ \mathbf{\Omega}(\lambda_1\mathbf{\Omega} + \lambda_2\mathbf{I}_M)^{-1}\mathbf{e}_{m2}(x_{i1}^{m1})^T x_{i2}^{m2}$, and $\mathbf{\Lambda}$ as a diagonal matrix with elements value N if the corresponding data point belong to the m-th attribute classifier.

Optimizing Ω when W and b are fixed

If **W** and **b** are fixed, the problem in Eq. (6) can be has an analytical solution

$$\mathbf{\Omega} = \frac{(\mathbf{W}^T\mathbf{W})^{\frac{1}{2}}}{tr((\mathbf{W}^T\mathbf{W})^{\frac{1}{2}})} \qquad (8)$$

The above two steps are performed alternatively, until the optimization procedure converges or the maximal iteration number is reached.

4 Incorporating Attribute Relation to Zero-Shot Learning

After we capture the attribute relation automatically in previous section, we want to use it explicitly for improving the zero-shot learning performance. Fig.1 illustrates the overall flowchart of our method. First of all, we use all the training data points and their attributes to train M attribute classifiers. And then, these classifiers are employed to predict the attribute values of unseen test images, and each test image will be given M attributes of real values. With specific inventory of attributes for each object class, predicted attributes will be mapped into class labels through DAP technique [10].

It's worth noting that the relation learned by our proposed method can be incorporated into the attribute-label prediction process easily. The underlying intuition is to join traditional dependent attribute classifiers together through our learned attribute relation. As is shown in Fig.1, Attribute value vector Y predicted by traditional classifiers, e.g. SVM and KRR, can be modified by attribute relation through $Y\_modi = Y * Cor$, where $Y \in \mathbb{R}^U$, U is the number of test images, and $Cor \in \mathbb{R}^{M \times M}$ is the attribute correlation matrix we learned from ARL method.

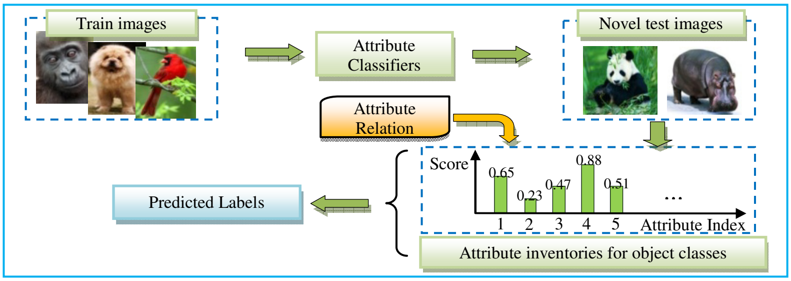


Fig. 1. Overview of our attribute-based zero-shot object classification

5 Experiments

5.1 Experimental Setup

We use a subset of the *Animals with Attributes* (*AWA*) dataset [14], which consists of 50 animal categories and 85 attributes. Following [10], only 59 attributes are employed in our experiments. For computational reasons, we down-sample this data set to 100 images per category in our experiments. Six feature types are used as image descriptors. And 10 classes are used as testing set and the other 40 classes as training set. We also use the *a-Pascal-train* data set [1] which consists of 20 classes and 64 attributes. As is given in [3], color and texture features are employed as image descriptors. Five classes (*bus, car, cat, dog and motorbike*) are used as test data while the other 15 ones as training data. We employ the χ^2-kernel [21] for image representation, i.e. kernel matrixes are based on χ^2-kernel of individual feature type.

Two baselines are employed, i.e. SVM and kernel ridge regression (KRR). Normalized mutual information (NormMI) used in [4,13] are employed as baseline for attribute relation learning. We adopt area under ROC curve (AUC) to evaluate the attribute prediction results. Average classification accuracies are used as performance measure for zero-shot learning methods. The regularization parameters λ_1 and λ_2 for our ARL method are both selected on from {0.001, 0.01, 0.05, 0.1, 0.5, 1}, and the parameter C for SVM and λ for KRR are chosen from {0.001, 0.01, 0.1, 1, 10, 100, 1000}. Optimal parameter values are confirmed on a validation set of training classes.

A sigmoid transform [22] maps the outputs of attribute classifiers into probabilistic scores for DAP model, where optimal values for parameter A and B are set on the same validation set.

5.2 Results and Discussion

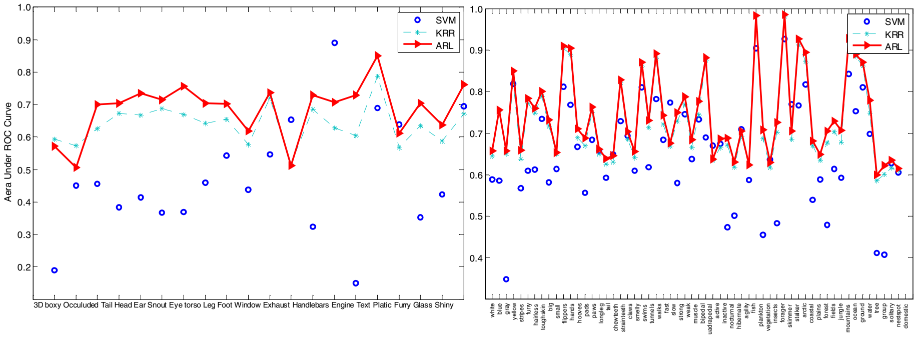
5.2.1 How about Classifiers Learned from ARL in Attribute Relation?

Next, we examine how the proposed ARL method performs in attribute prediction, with comparison to SVM and KRR. The quality of the individual attribute predictors is shown in Fig.3. And Table1 reports the AUC values of each attribute.

Table 1. Average AUC of attribute prediction on two data sets (%)

Data Sets	SVM	KRR	ARL
aPascal	47.2	63.8	**68.5**
AWA	65.3	72.9	**74.4**

From Fig.3, it's obvious to find that our ARL method outperforms SVM and KRR on average. Due to the fact that testing classes are different from the training ones, the across category generalization of the attribute classifiers learned from our ARL method is quite reliable. From Table1, we can see a significant improvement in average accuracy is obtained by ARL model comparing to SVM and KRR.



(a) aPascal (b) AWA

Fig. 2. Attribute prediction Results on *aPascal* and *AWA* data sets

5.2.2 How Does the Attribute Relation Influence Zero-Shot Learning Accuracy?

We now study the influence of attribute relation on the performance of zero-shot learning problem. Attribute relation learned from ARL method is used to modify the output of SVM and KRR. Table 2 reports the average accuracy of test classes in *AWA* data set. Note the term "*+*Corr*" means all correlation coefficients (positive, negative and uncorrelated) are incorporated, while the "*+*posCorr*" means only the positive ones are employed. From the results inTable 2, we draw the following conclusions.

First, our ARL model is superior to SVM and KRR in most cases, with the highest accuracy of 17.3% using Phog features. As is expected, ARL considers the attribute relation during the procedure of training attribute classifiers while SVM and KRR not.

Second, KRR and SVM with attribute relation are generally superior to those without correlation. It further validates our intuition that mining the associations between attributes helps improve performance of zero-shot object classification tasks.

Table 2. Zero-shot performance of different methods on *AWA* data set (%)

Method	Cq	Lss	Sift	rgbSift	Phog	Surf	All features
SVM	15.4	15.4	15.4	14.3	14.0	15.4	13.0
SVM +posCorr	**20.7**	13.9	14.5	**22.3**	**22.2**	6.7	13.8
SVM +Corr	12.9	9.2	**20.9**	7.6	16.1	9.1	9.2
KRR	12.0	10.0	5.6	13.9	9.7	10.5	12.8
KRR+posCorr	17.2	10.6	15.2	16.5	13.4	7.5	18.1
KRR+Corr	13.8	10.1	18.7	13.3	9.8	**17.8**	13.5
ARL	15.4	**17.0**	14.5	15.4	17.3	10.0	**15.7**

6 Conclusion

In this paper, we focus on attribute relation learning for attribute-based zero-shot object classification. We first propose an attribute relation learning (ARL) method to learn the correlation between attributes explicitly. Then, we present an attribute-based zero-shot object classification scheme with attribute relation incorporated. Finally, we further study how the attribute relation learning help improve the performance of zero-shot learning. Experimental results have validated the effectiveness of our proposed method.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under grant Nos. 60875030, 60973097 and 61072148, and by the Specialized Research Fund for the Doctoral Program of Higher Education under grant No. 20123218110009.

References

1. Farhadi, A., et al.: Describing Objects by Their Attributes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2009)
2. Rohrbach, M., et al.: What Helps Where - and Why? Semantic Relatedness for Knowledge transfer. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2010)
3. Lang, H., Ling, H.: Classifying Covert Photographs. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2012)
4. Wang, Y., Mori, G.: A Discriminative Latent Model of Object Classes and Attributes. Perspectives in Neural Computing, 155–168 (2010)

5. Siddiquie, B., et al.: Image Ranking and Retrieval based on Multi-attribute Queries. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2011)
6. Rohrbach, M., et al.: Evaluating Knowledge Transfer and Zero-Shot Learning in a Large-Scale Setting. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2011)
7. Larochelle, H., et al.: Zero-data Learning of New Tasks. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence (2008)
8. Mitchell, T.M., et al.: Predicting Human Brain Activity Associated with the Meanings of Nouns. Science 320, 1191–1195 (2008)
9. Palatucci, M., et al.: Zero-Shot Learning with Semantic Output Codes. In: Advances in Neural Information Processing Systems (2009)
10. Lampert, C.H., et al.: Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2009)
11. Farhadi, A., et al.: Attribute-centric Recognition for Cross-category Generalization. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2010)
12. Parikh, D., Grauman, K.: Interactively Building a Discriminative Vocabulary of Nameable Attributes. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2011)
13. Kovashka, A., et al.: Actively Selecting Annotations among Objects and Attributes. In: IEEE International Conference on Computer Vision (2011)
14. Argyriou, A., et al.: Convex Multi-task Feature Learning. Machine Learning 73, 243–272 (2008)
15. Bonilla, E., et al.: Multi-task Gaussian Process Prediction. In: Proceedings of NIPS (2007)
16. Zhang, Y., Yeung, D.-Y.: A Convex Formulation for Learning Task Relationships in Multi-Task Learning. In: Proceedings of UAI, pp. 733–742 (2010)
17. Rukhin, A.L.: Matrix Variate Distributions. Journal of the American Statistical Association 98, 462, 495–496 (2003)
18. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
19. Van Gestel, T., et al.: Benchmarking Least Squares Support Vector Machine Classifiers. Machine Learning 54, 5–32 (2004)
20. Keerthi, S.S., Shevade, S.K.: SMO Algorithm for Least-squares SVM Formulation. Neural Computation 15, 487–507 (2003)
21. Teytaud, O., Jalam, R.: Kernel-based Text Categorization. In: International Joint Conference on Neural Networks, vol. 3, pp. 1891–1896 (2001)
22. Platt, J.C.: Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods. In: Smola, A.J., Bartlett, P., Schölkopf, B., Schuurmans, D. (eds.) Advances in Large Margin Classifiers, pp. 61–74 (1999)