Multimed Tools Appl (2016) 75:7029–7043
DOI 10.1007/s11042-015-2630-5

Image attribute learning with ontology guided fused lasso

Chao Li[1] · Zhiyong Feng[1] · Yahong Han[1,2]

Received: 21 January 2014 / Revised: 15 March 2015 / Accepted: 13 April 2015 /
Published online: 21 May 2015

Abstract Extended from the traditional pure statistical learning methods, we propose to augment the statistical learning methods with ontology and apply this idea for image attribute learning. In order to capture structural information among attributes, the graph-guided fused lasso model is adopted and improved by a new distance metric based on WordNet. The novelty of our method is that we find the semantic correlation with the ontology-guided attribute space and integrate inter-attribute similarity information into the learning model. The hierarchy of ImageNet is exploited to define the image attributes and a dataset from ImageNet including over 30,000 images is collected. The experimental results show that this method can both improve the accuracy and accelerate the algorithm convergency. Moreover, the learned semantic correlation owns transfer ability to related applications.

Keywords Image attribute learning · Ontology · Graph-guided fused lasso · Transfer learning

1 Introduction

With the big-data era approaching, the large-scale web images bring out a great challenge to image understanding and retrieval. Thus, related works like image classification and automatic image annotation have been well explored.

✉ Yahong Han
yahong@tju.edu.cn

Chao Li
qizuma@tju.edu.cn

Zhiyong Feng
zyfeng@tju.edu.cn

[1] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
[2] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China

Previous works on automatic image annotation mainly focus on the better probabilistic representations and the adoption of learning-based methods [10, 14, 19]. However, the prior or domain knowledge has been ignored. Knowledge is used by humans when learning the visual appearance of objects [15]. For example, babies sometimes learn new things by the knowledge their parents or teachers tell them. So in our opinion, knowledge is useful for learning and thus ontologies are particularly important.

In this paper, we propose to augment the statistical learning methods with ontology and apply this idea to image attribute learning. The so-called attributes are interpreted as inherent characteristic in Webster dictionary, which are intrinsic human-nameable qualities of images. Attribute-based ideas have been shown to be helpful in various applications like face verification [18], image retrieval [28, 37], action recognition [21], robotics and mobile communications [16], and zero-shot transfer learning [16, 19, 25, 36, 38].

However, knowledge embedded in the inter-attribute relationship is rarely considered and human efforts are usually involved such as to label the attributes. To solve these issues we propose a method called *Image Attribute Learning with Ontology Guided Fused Lasso (IAL-OGFL)*. Ontologies are used for mining inter-attribute similarity and graph-guided fused lasso [17] is exploited for sparse feature selection.

Why Ontology? knowledge representation is central to the applications of knowledge-based methods. According to the "modelling view" of knowledge acquisition proposed by Clancey [4], a knowledge base is not a repository of knowledge extracted from one expert's mind, but the result of a modeling activity whose object is the observed behavior of an intelligent agent embedded in an external environment. This implies that it may not get good results for learning by exploring experiential knowledge to some extent. For example, many papers acquire priors from a manual class-attribute correlation matrix for attribute learning recently [13]. Since the matrix is generated with some skilled workers, which is not authoritative and hard to reuse when changing the labels of classes and attributes, its suitable to be improved with ontology. From another perspective, according to the kinds of primitives used, knowledge representation formalisms can be classified into five categories (Fig. 1) [12]. We can see from Fig. 1 that interpretation with logical and epistemological is arbitrary and with conceptual and linguistic is subjective. But in the ontological level, the ontological commitments associated to the language primitives are specified explicitly which can restrict the number of possible interpretations. For these characteristics and the purpose of sharing and reuse of knowledge, we propose to utilize ontology in image attribute learning.

Why Graph-Guided Fused Lasso? For the images in the real world, high dimensional low-level features can be extracted and only a small fraction of them are associated with

Level	Primitives	Interpretation	Main feature
Logical	Predicates, functions	Arbitrary	Formalization
Epistemological	Structuring relations	Arbitrary	Structure
Ontological	***Ontological relations***	***Constrained***	***Meaning***
Conceptual	Conceptual relations	Subjective	Conceptualization
Linguistic	Linguistic terms	Subjective	Language dependency

Fig. 1 Classification of knowledge representation formalisms

their corresponding attributes. So it may increase the computational complexity without feature selection. Lasso [31] is suitable for sparse feature selection, but it is incapable of capturing any structural information among attributes, structured-sparsity-inducing penalty should been considered [3]. Unlike the group lasso separating attributes into groups and fused lasso treating attributes as chain structure, graph-guided fused lasso introduced a general class of structure and therefore more priors can be included.

The main contributions of our work can be summarized as follows:

1) Inter-attribute similarity is integrated into the graph-guided fused lasso model. Different from previous works, the WordNet-based metric space is exploited for inter-attribute similarity measurement (Section 3).
2) The idea that statistical learning is directed with ontology is shared and a principled framework of *IAL-OGFL* is proposed (Section 4).
3) Comprehensive experiments are conducted to demonstrate the effectiveness of our approach. Our method has outstanding performance with higher accuracy rate and faster convergence than similar works (Section 5).

2 Related work

Attribute-based methods have received much attention in the area of computer vision. Ferrari et al. [10] and Lampert et al. [19] presented a series of interesting applications which have demonstrated the power of semantic attributes. The probabilistic generative model [10] and the Direct Attribute Prediction (DAP) model [19] both treats each visual attribute as independent and train the attribute classifiers not considering their relationships. For example, the DAP model [19] trains a non-linear support vector machine (SVM) for each binary attributes and no inter-attribute information exchange in this process.

However, in the real world, dependencies between attribute pairs are ubiquitous which has also been proved by [13] with Animal with Attributes (AwA) database. For example, "ocean" has strong correlation with "water" and a weak correlation with "dessert" in AwA. Many methods considering the dependencies have been proposed. Hwang et al. [14] believed that all attributes can rely on some shared structure in the low level feature space, so a convex multi-task feature learning method with an ℓ_1/ℓ_2-norm is adopted. But according to the research of [13], some attributes are more likely to share common relevant low-level features, and they proposed a method with graph-guided fused lasso which exploits graph to describe the correlations of attributes. Similarly, Yu et al. [37] design a novel two-layer probabilistic graphical model for finding the relevance of attributes. Wang et al. [35] also proposed a discriminative model for joint modeling object class labels and their attributes. They also assumed there are certain dependencies between some attribute pairs and an attribute relation graph is used for their model. Zhang et al. [39] proposed a method to organize the semantic concepts into multiple semantic levels and argument each concept with a set of related attributes. Their method is used for image retrieval and achieves good results.

The common point of the above graph-based methods is that they explore experts experiential knowledge for learning. For example, Han et al. [13] constructed the graph with a manual class-attribute correlation matrix by skilled workers. The matrix is illustrated to be intuitive but not discriminative possibly [16, 22, 26, 36].

Instead, we share the idea that statistical learning can be directed with ontology. An ontology formally represents knowledge as a set of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts [11]. Comparing with experts experiential knowledge, an ontology is a more formal representation of a set of concepts and their relationship, so it is more authoritative for mining inter-attribute similarity information. Whats more, inter-attribute similarity information can be pre-learned easily with ontologies no matter how attributes scales. Ontologies have been widely used for designing concepts correlations in the area of computer vision such as image annotation [7, 27, 29, 33], object detection [1, 2, 6, 23, 30], image retrieval [24, 34] and scene understanding [20]. For example, a concept ontology composed of several types of concepts (spatial concepts and relations, color concepts and texture) was combined with machine learning techniques, which was used for complex object recognition in [6]. The strength of this method is that the visual concept ontology acts as user-friendly intermediate between image processing layer and the expert. Li et al. [20] proposed a hierarchical generative model for scene classification, object component segments, and image annotation. WordNet was used in order to provide a handful of relatively clean images in which some object regions are marked with their corresponding tags. In order to solve the problem that the returned results of ranking methods for tag-based image search are irrelevant or not diverse, Wang et al. [34] proposed a diverse relevance ranking scheme, in which WordNet is used for words relevance estimation.

3 Ontology guided fused lasso

The Ontology Guided Fused Lasso (*OGFL*) is a model proposed to compute the relevancy between features and attributes. In this section, we first present a definition of the proposed Ontology Guided Fused Lasso as follows:

Definition 1 Ontology Guided Fused Lasso is defined as $OGFL = (I, T, M)$.

Initial State: $I = \{X^S, Y^S, G\}$stands for the initial state the model, where $X^S \in R^{N\times P}$ represents the source image data matrix for N samples and P-dimensional features and $Y^S \in \{0, 1\}^{N\times L}$ is the attribute indicator matrix of source image data for L attributes. G is an inter-attribute similarity graph constructed with ontologies.

Terminal State: $T = \{B\}$ stands for the terminal state of the model. B is the feature-attribute relevancy graph represented with matrix, where each column is a P-vector of regression coefficients for every attribute.

Model: The model used to bridge from the initial state I to the terminal state T is graph-guided fused lasso $(\min_B \parallel Y^S - X^S B \parallel + \gamma \Omega_G (B) + \lambda \parallel B\|_1)$.

As ontologies can mine the inter-attribute semantic similarity and graph-guide fused lasso leads attributes to be more similar in the low-level feature space, *OGFL* which integrates ontologies with the graph-guided fused lasso can bridge the low level feature space with the high level attribute space naturally. Hence, the learned feature-attribute graph will be a convenient model for selecting the most valuable features for every attribute and attribute learning.

In this section, *OGFL* will be introduced in detail. We will first introduce how to combine ontology with the graph-guide fused lasso model (Section 3.1).Then we will describe the construction of the ontology guided inter-attribute similarity graph, which is built in the WordNet-based metric space (Section 3.2).

3.1 Graph-guided fused lasso with ontology

Assume that we have a set of L attributes for the problem of attribute learning. Lasso tends to solve a set of L independent regressions for each attribute with its own L1 penalty, and it doesn't provide a mechanism to combine information across multiple attributes such that the similarity can be reflected in the regression coefficients for those correlated attributes. However, several attributes are often highly correlated and they often share some structures in the feature space. That is to say, highly correlated attributes may share more features. So it is difficult for lasso to describe this characteristic.

GFlasso extends the standard lasso, and it is a new penalized regression method with pleiotropic effect on correlated attributes. *GFlasso* regards the correlation structure over the set of L attributes as an edge-weighted graph, and use this graph to guide the learning process. The *GFlasso* is particularly suitable for attribute learning problems because no attribute is isolated and universal correlation exits between attributes. As mentioned above, the *GFlasso* model we used in the attribute learning problem is:

$$\min_{B} \parallel Y^S - X^S B \parallel + \gamma \Omega_G (B) + \lambda \parallel B \parallel_1 \tag{1}$$

Where $Y \in \{0, 1\}^{N \times L}$ is the attribute indicator matrix of source image data. $X \in \mathbf{R}^{N \times P}$ represents the source image data matrix and B is the mapping of the feature space and attribute space trying to get. γ and λ are regularization parameters respectively that control the complexity of the model. A larger value of γ leads to a greater fusion effect.Considering that effective features for every attribute are usually sparse, regular Lasso ($\parallel B \parallel_1 = \sum \sum \parallel B(:,:) \parallel$) is used. However, Lasso is prone to selecting features individually. As described above, attributes share some structures in the feature space. That is to say, highly correlated attributes may share more features, which is beneficial semantics for attribute feature selection. In order to encode the structured priors of attribute correlation into the model, graph penalty $\Omega_G (B)$ is considered:

$$\Omega_G (B) = \sum_{e=(m,l) \in E, m<l} \tau (r_{ml}) \, | B_m - sign(r_{ml}) B_l | \tag{2}$$

Where B_m and B_l are the m_{th} and l_{th} columns of B respectively and they are the regression coefficients for the m_{th} and l_{th} attributes. $\tau (r) = |r|$ weights the fusion penalty for each edge of graph G. $sign(r_{ml}) = 1$ for two positively correlated attributes and $sign(r_{ml}) = -1$ for two negatively correlated attributes. $\Omega_G (B)$ encourages B_m and B_l to take the same value by shrinking the difference between them toward 0.

Assume that we have construct an ontology guided inter-attribute correlation graph G_o from a preprocessing step consisting of a set of nodes V, each representing one of the L attributes and a set of edges E. The weight of each edge $(m, l) \in E$ is sim_{ml} standing for the relevancy of every two attributes. For two high correlated attributes m and l in the feature space (low value of $|B_m - sign(r_{ml}) B_l|$), they should very close in the attribute space (high value of sim_{ml}). Hence, $\tau (r_{ml})$ can be replaced with sim_{ml} in (2), which enriches the interpretability and can improve the accuracy of attribute learning as shown in the experiment part.

3.2 WordNet-based metric space and attribute relation graph construction

Considering the information of inter-attribute similarity, there are some ways to construct the graph for attribute learning problem. In [13] a class-attribute matrix which is constructed with skilled workers is exploited for clustering in order to get an inter-attribute similarity

graph A(see in Fig. 2), and [17] adopts an approach which computes pairwise Pearson correlation coefficients for all pairs of attributes using the label matrix Y. These methods are statistical, and the time complexity will increase with the increasing number of classes of A and the growing numbers of samples of Y, which makes them to have poor expansibility. Besides, experts experiential knowledge is required for labeling the attributes for every class in [13], it is less objective and authoritative than knowledge extracted from ontologies. We proposed a method constructing attribute graph with ontology and without learning, which is simple and effective.

We construct graphs with WordNet. Information in WordNet is grouped into sets of cognitive synonyms (*synsets*). *Synsets* are interlinked by means of conceptual-semantic and lexical relations. We adopt a simple and commonly used approach for learning such graphs in this article, where we first compute pairwise *WUP* similarity (Wu and Palmer, 1994) for all pairs of attribute in WordNet, and then connect every two nodes with an edge to build the graph.

WUP views WordNet as a graph and is a function of the path length from the lowest super-ordinate (*LSO*) of the two concepts *m* and *l*, which is the most specific concept that they share as an ancestor. For example, if *m* was '*pest#n#4*' and *l* was '*arthropod#n#1*'

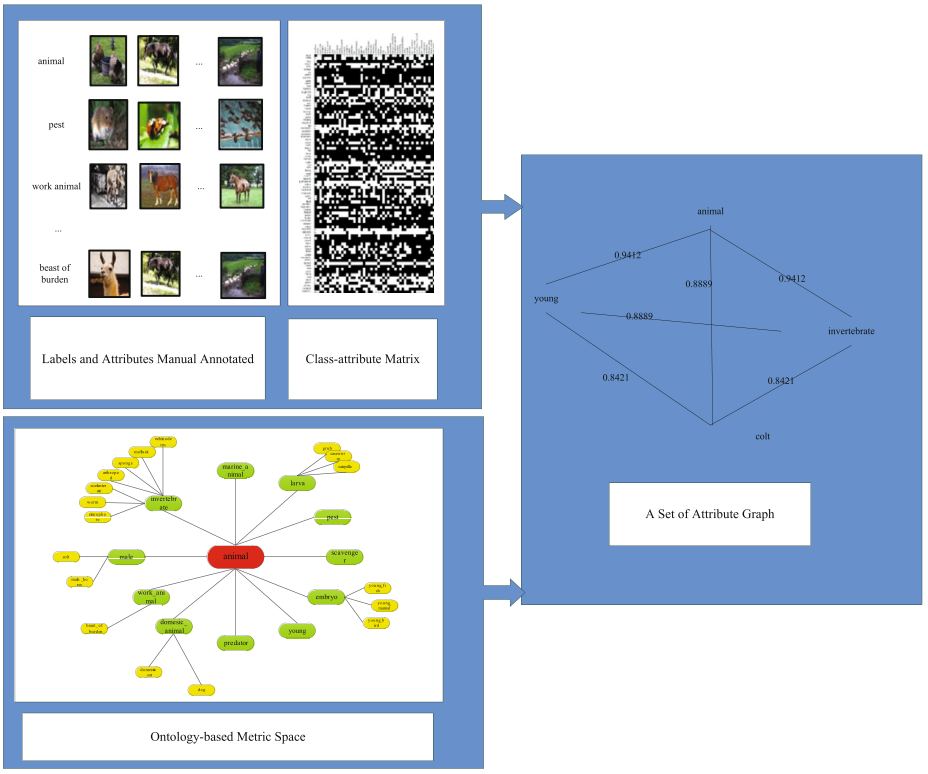


Fig. 2 Attribute graph construction with learning-based (*upper left*) ideas and ontology-based ideas (*lower left*). In order to acquire inter-attribute similarity information, clustering strategy is applied to learning-based idea with a manual labeled class-attribute matrix while human prior knowledge is obtained with ontology-based ideas

then the $LSO\ (m, l)$ would be '*animal#n#1*'. The *WUP* similarity between m and l can be calculated as follows:

$$sim_{ml} = \frac{2 \times depth\ (LSO\ (m, l))}{len\ (m, LSO\ (m, l)) + len\ (l, LSO\ (m, l)) + 2 \times depth\ (LSO\ (m, l))} \tag{3}$$

Where $len\ (m, LSO\ (m, l))$ measures the length of the shortest path in WordNet from concept m to concept $LSO\ (m, l)$, $depth\ (LSO\ (m, l))$ means the length of the path to $LSO\ (m, l)$ from the global root, i.e. $depth\ (LSO\ (m, l)) = len\ (root, LSO\ (m, l))$.

The semantic relations between attribute *'pest#n#4'* and attribute *'arthropod#n#1'* can be calculated as in Fig. 3. The similarity between them is 0.8421 which means two concepts are closed enough. The *WUP* measurement is simple with low complexity. It only relies

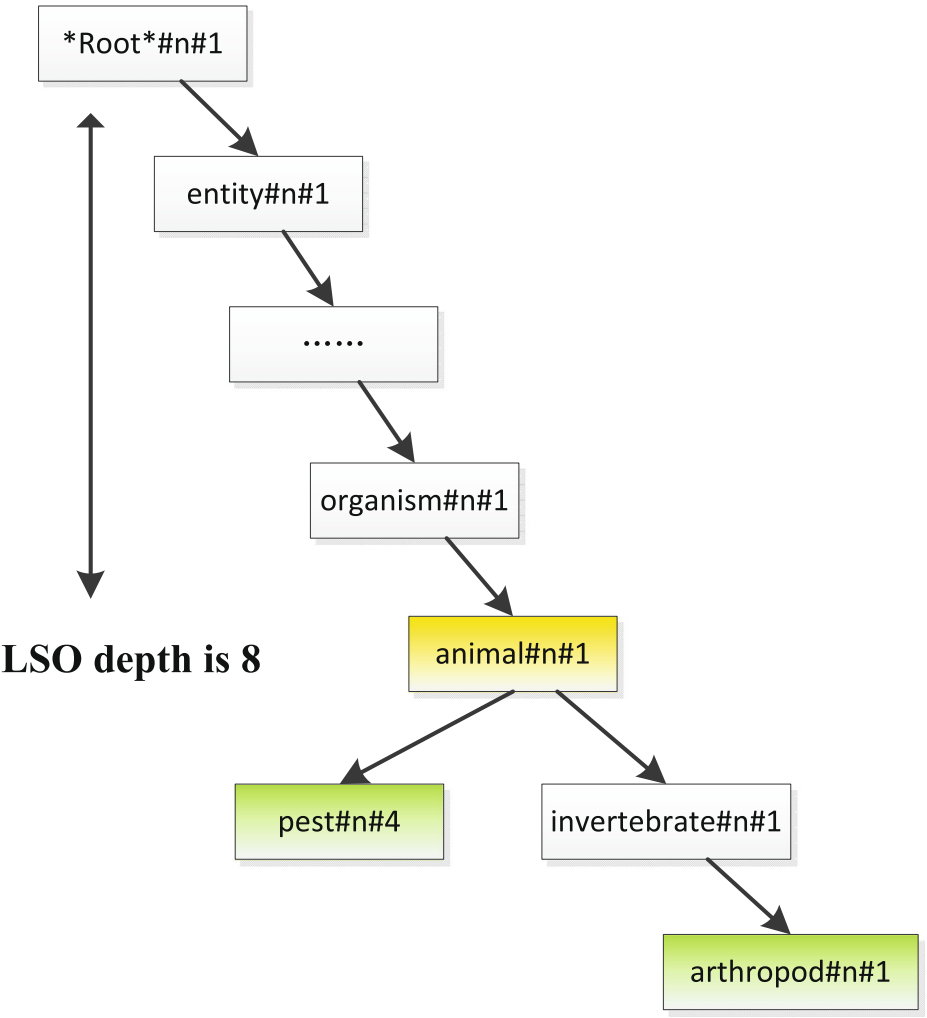


Fig. 3 WUP measurement of *pest#n#4* and *arthropod#n#1*. The *LSO* of the two attributes is *animal#n#1* and the depth of *animal#n#1* is the path from root to itself which is 8. Hence, the similarity between the two attribute is 0.8421

on the depth based on ontologies for every pair of attributes, and the complexity doesnt increase with the growing numbers of samples or classes. Besides, since ontologies is built in line with cognitive science, ontology guided learning gets better results.

4 Image attribute transfer with ontology

The learned matrix B with OGFL in Section 3 is integrated with inter-attribute similarities and corresponds to coupling pairs of attributes in the adjacent rows of the same column. Besides, it reflects the correlativity of every attributes with its features. A Larger value of element in B means a greater relevancy for the attribute with its corresponding feature. Hence, the matrix B is consistent with the assumptions mentioned before that there is a shared structure between the attribute space and the original image descriptor space, and it is very suitable for feature selection for individual attribute. Since the matrix B is learned with ontology, it reflects the intrinsic characteristics of attributes and is relatively easier to transfer learning with different samples or different databases.

Assume that we have a target image dataset $T = \{X^T\}$ with $X^T \in R^{N\times P}$ which can be annotated with the L attributes. Then an algorithm for feature selection and attribute transfer can be get (Algorithm 1). Every column of matrix B (e.g. $B_{(:,l)}$ l=1,,L) corresponds to one attribute and reflects how all the features influence the attribute. Hence, the characteristics of matrix B can be exploited to perform feature selection of every attribute for target images. We rank elements in vector $B_{(:,l)}$ according to the value of $\left\|B_{(p,l)}\right\|$ $(p = 1, ..., P)$ in descending order, and the top f features are the most beneficial features for $B_{(:,l)}$. After feature selection, various classifiers can be trained. In this paper, we have tried the knn classifier and SVM to test the result of feature selection. In this process, the correlated information among attributes is transferred from the source images to the target images in order to get a better representation for the target images.

Algorithm 1 Image Attribute Transfer Algorithm

Input:
feature-attribute coefficient matrix B; target images $T = X^T$; f is the number of features to be selected.

Output:
attribute indicator matrix Y^T for testing images of target data

1: For l =1,...,L
2: Rank elements in vector $B_{(:,l)}$ according to the value of $\left\|B_{(p,l)}\right\|$ $(p = 1, ..., P)$ in descending order;
3: Output the index ind_l of the top f selected features;
4: Select features for target images T with ind_l;
5: Training and testing the attribute prediction classifier;
6: End For
7: Output the attribute indicator matrix Y^T for testing images of target data.

The framework of *IAL-OGFL* can be illustrated with Fig. 4. The key points are as follows: (1) constructing a WordNet-path-based metric space and mining semantic relation of attributes to construct the graph (Section 3.2); (2) using the pre-learned inter-attribute correlation graph and source samples to solve the graph-guided fused lasso model with a smoothing proximal gradient method proposed in [3] with multi-task extension (For reasons of space, it is not introduced here)(Section 3.1); (3) transferring the matrix to selecting features of every attribute with target samples; (4) predicting attributes with the selected visual features.

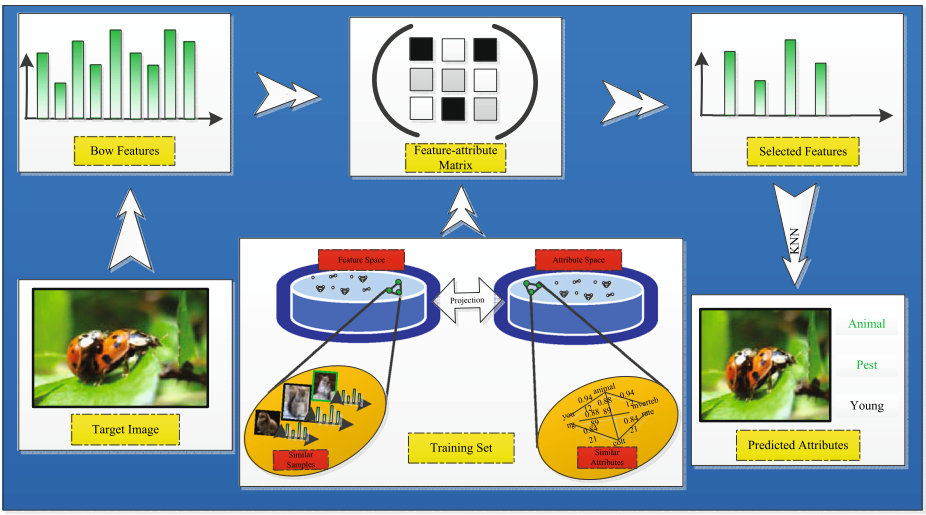


Fig. 4 The framework of *IAL-OGFL*. For the training set, a projection from the low-level feature space to the ontology-based attribute space is found which encourages high correlated attributes sharing similar features. The projection is represented with a feature-attribute correlation matrix (top center) and can be used for feature selection and attributes prediction for the target image set

5 Experiment and result

5.1 Dataset and image features

Generally speaking, attributes are usually designed by manually picking a set of words. In [8], besides semantic attributes, discriminative attributes (e.g. "cats and dogs have it but sheep and horses dont") are designed by experts. In [19], attributes are collected by experts according to "relative strength of association" between attributes and classes. The common ground of these methods is that additional human efforts are involved. To solve this problem, Yu et al. [36] proposed to design "category-level attributes" which will not have concise names as the manually specified attributes. Unlike these methods, the hierarchical of ImageNet is taken advantage of to acquire attributes which we think is easy and suggestive.

Semantic hierarchies are always used for image annotation [32]. Similarly, in this paper, the hierarchy of ImageNet is exploited to define the image attributes. The hierarchy of ImageNet is built mostly cording to hyponymy, which is also called "is-a" relation. For example, a "human" is an "animal", and a "worm" is an "invertebrate". The "is-a" relation is a very important inherent characteristic. Naturally, we treat the father node as the attribute of the son node. For example, "animal" can be exploited as an attribute of "male" and "invertebrate" is an attribute of "worm".

ImageNet contains over 10 million images and over 15000 synsets (sets of cognitive synonyms)[5]. We do our experiment on the animal branch. 30 classes (see Fig. 5) in 3 layers with 31288 images are selected to build the dataset. The number of images in each class is various, ranging from tens of pictures to thousands of pictures.

SIFT Bag of Visual Words feature is used in our experiment for its robustness with image rotation and stability with visual angle variation. First SIFT (Scale-Invariant Feature Transform) points are extracted for the entire image in the database. Then the randomly selected

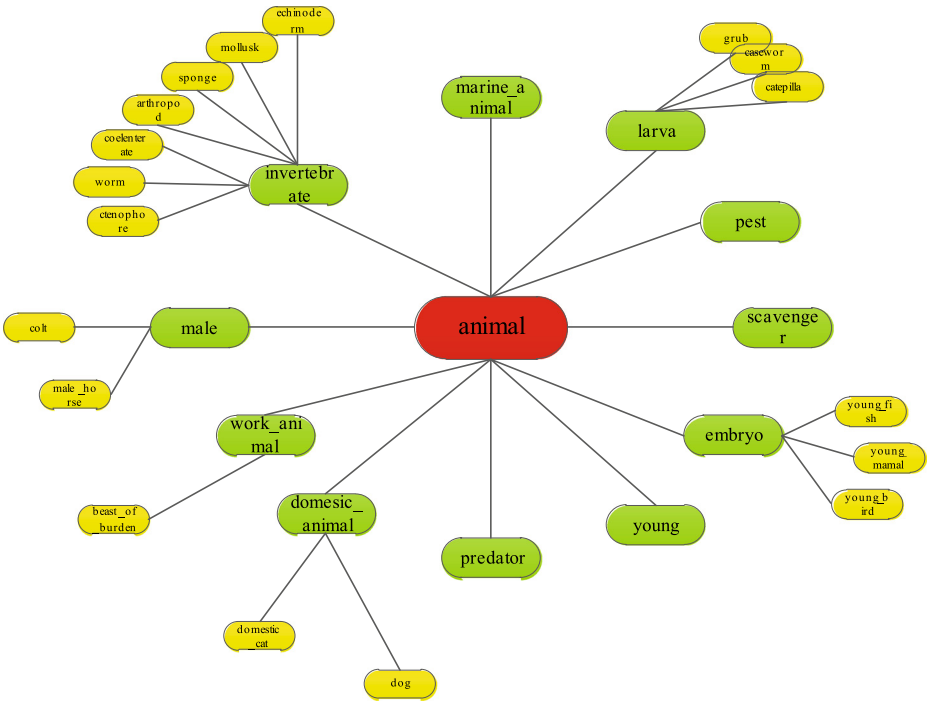


Fig. 5 Attributes our experiment used and their hierarchy. All these attributes are from ImageNet

set of SIFT points are clustered and produced the 1,000 centers as the visual dictionary. At last each image is quantized into a 1,000 dimensional histogram of bag-of-visual-words.

5.2 Parameter tuning

In order to train a optimal regressor with (2), the weight value λ and γ need to be determined first. Since B is trained and learned from regression process, the values of predicted responses are continuous but binary. Thus, the Area under the Roc Curve (AUC) [9] is used as the evaluation metric. We use different parameters of λ and γ and select the best. The ranges of λ and γ are {0.0001, 0.001, 0.01, 0.1, 1, 10, 100 }. For every value of λ and γ, we randomly select half images for training and remaining for testing, each experiment is repeated for 10 times and the average value is reported as in Table 1. The highest AUC value 0.780219 indicates the best result when λ=γ=1.

5.3 Ontology guided fused lasso performance

A source image set containing randomly selected 22207 images with 30 attributes is used for training and testing the performance of the Ontology Guided Fused Lasso model. We compare our method with CAT-MtG2F (correlated attribute transfer with multi-task graph-guided fusion) [13] and other flat methods.

As shown in Table 2, with a no-graph method that only uses ℓ_1 norm (ℓ_1 Method), a flat graph-guided fusion method where all attributes have the correction of 1 with the other attributes (FlatMtG2F), a kNN based CAT-MtG2F method with different k (kNN-MtG2F),

Table 1 AUC value with different parameters λ and γ

γ \ λ	0.0001	0.001	0.01	0.1	1	10	100
0.0001	0.745065	0.745069	0.745112	0.745528	0.749555	0.771694	0.681005
0.001	0.745069	0.745074	0.745115	0.745533	0.749558	0.771697	0.681007
0.01	0.745383	0.745387	0.74543	0.74584	0.749842	0.771817	0.681014
0.1	0.752042	0.752046	0.752081	0.752459	0.75615	0.774625	0.680663
1	0.777552	0.777555	0.777588	0.777904	0.780219	0.77721	0.673631
10	0.621223	0.621226	0.621261	0.621678	0.624428	0.664172	0.669373
100	0.720758	0.720759	0.720769	0.720868	0.721958	0.734836	0.683619

a Pearson correlation coefficient based method (PearsonCC) used in [17] and our ontology-based idea, we randomly select half images from the source image set for training and left for testing, each experiment is repeated for 10 times. The average iterations, running time and AUC value of B are reported.

It shows in Table 2 that our ontology guided method is much easier to convergence with the least iterations and has the highest AUC value. That means that the WordNet-based metric space is more like to describe the inter-attribute similarity. Hence, the ontology guided regressor is more discriminative. Our method uses less time compare with FlatMtG2F and CAT-MtG2F when $k \neq 1$ which also implies that the graph constructed with ontology is better.While our method is slower than ℓ_1 Method and 1NN-MtG2F because their methods are simpler with fewer constraints (ℓ_1 Method has no graph and 1NN-MtG2F has a simple graph with attributes having no corrections with the others).

5.4 Result of attribute transfer

We use classifiers to verify the effectiveness of the learned matrix B. The remaining 10000 images with 30 attributes are used as the target image set. We randomly select 90 % of the samples as training set and the remaining for test. Feature selection is performed on every attribute with 50 features selected, and every attribute is attached with a classifier.

We test the learned model with SVM classifier. We use libSVM and the best C and G is selected for every classifier. Our method is compared with CAT- MtG2F, PearsonCC and PCA by accuracy and mean squared error (Table 3). From Table 3 we can see our method also has the best performance with accuracy and mean squared error.

Table 2 Performance of B with different methods

Method \ Evaluation metrics	Iterations	Time	AUC value
ℓ_1 Method	93.7	4.204389	0.749555
FlatMtG2F	84.9	4.55791	0.777337
1NN-MtG2F	93.7	3.797617	0.749555
10NN- MtG2F	93.4	4.435152	0.766852
30NN- MtG2F	89.1	4.724376	0.768709
PearsonCC	91	5.158014	0.766890
Our Method	83.9	4.344154	0.780219

Table 3 Accuracy and Mean Squared Error with libSVM

	Accuracy	Mean Squared Error
OurMethod	94.77241	0.040721
CAT-MtG2F	94.72759	0.056206
PearsonCC	94.67931	0.053207
PCA	94.71725	0.051414

6 Conclusion

We have augmented the statistical learning methods with ontology and proposed a novel ontology guided fused lasso method for image attribute learning. Our method has several advantages compared with previous methods. Firstly, we obtain the priors of interrelationship of attributes from ontology, which is more explicable relative to pure statistical methods. Secondly, a WordNet-path-based metric is used for designing inter-attribute correlations, which is very flexible, which can be easily modified to improve upon many different performance measurements. Thirdly, the WordNet-based attribute space has the advantage to scale up the process to develop a large number of attributes. The experiments show that our method can both accelerate the convergence and improve the accuracy rate with SVM classifier. It implies that the WordNet-based metric space is more like to describe the inter-attribute similarity and proves that ontology is beneficial for learning. As well, the Ontology Guided Fused Lasso has outstanding transfer ability. Our future work is to consider various measurements with different ontologies and find a more feasible metric universally.

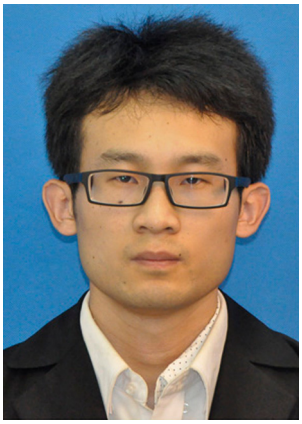
Acknowledgement This work was partly supported by the NSFC (under Grant 61202166, 61472276) and Doctoral Fund of Ministry of Education of China (under Grant 20120032120042).

References

1. Benitez AB, Chang SF (2003) Image classification using multimedia knowledge networks. In: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol 3, pp III–613–16 vol 2. IEEE
2. Breen C, Khan L, Ponnusamy A (2002) Image classification using neural networks and ontologies. In: Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on, pp 98–102. IEEE
3. Chen X, Lin Q, Kim S, Carbonell JG, Xing EP (2012) Smoothing proximal gradient method for general structured sparse regression. Ann Appl Stat 6(2):719–752
4. Clancey WJ (1993) The knowledge level reinterpreted: Modeling sociotechnical systems. Int J Intell Syst 8(1):33–49
5. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 248–255. IEEE
6. Eric Maillot N, Thonnat M (2008) Ontology based complex object recognition. Image Vis Comput 26(1):102–113
7. Fan J, Gao Y, Luo H (2008) Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. IEEE Trans Image Process 17(3):407–426
8. Farhadi A, Endres I, Hoiem D, Forsyth D (2009) Describing objects by their attributes. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 1778–1785. IEEE
9. Fawcett T (2006) An introduction to roc analysis. Pattern Recogn Lett 27(8):861–874
10. Ferrari V, Zisserman A (2007) Learning visual attributes. In: Advances in Neural Information Processing Systems, pp 433–440
11. Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5(2):199–220

12. Guarino N (1995) Formal ontology, conceptual analysis and knowledge representation. Int J Hum Comput Stud 43(5):625–640
13. Han Y, Wu F, Lu X, Tian Q, Zhuang Y, Luo J (2012) Correlated attribute transfer with multi-task graph-guided fusion. In: Proceedings of the 20th ACM international conference on Multimedia, pp 529–538. ACM
14. Hwang SJ, Sha F, Grauman K (2011) Sharing features between objects and their attributes. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp 1761–1768. IEEE
15. Jolicoeur P, Gluck MA, Kosslyn SM (1984) Pictures and names: Making the connection. Cogn Psychol 16(2):243–275
16. Kankuekul P, Kawewong A, Tangruamsub S, Hasegawa O (2012) Online incremental attribute-based zero-shot learning. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp 3657–3664. IEEE
17. Kim S, Sohn KA, Xing EP (2009) A multivariate regression approach to association analysis of a quantitative trait network. Bioinformatics 25(12):i204—i212
18. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In: Computer Vision, 2009 IEEE 12th International Conference on, pp 365–372. IEEE
19. Lampert CH, Nickisch H, Harmeling S (2009) Learning to detect unseen object classes by between-class attribute transfer. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 951–958. IEEE
20. Li LJ, Socher R, Fei-Fei L (2009) Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 2036–2043. IEEE
21. Liu J, Kuipers B, Savarese S (2011) Recognizing human actions by attributes. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp 3337–3344. IEEE
22. Mahajan D, Sellamanickam S, Nair V (2011) A joint learning framework for attribute models and object descriptions. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp 1227–1234. IEEE
23. Marszalek M, Schmid C (2007) Semantic hierarchies for visual object recognition. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pp 1–7. IEEE
24. Mezaris V, Kompatsiaris I, Strintzis MG (2003) An ontology approach to object-based image retrieval. In: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol 2, pp II–511–14 vol 3. IEEE
25. Russakovsky O, Fei-Fei L (2012) Attribute learning in large-scale datasets. In: Trends and Topics in Computer Vision, pp 1–14. Springer
26. Sharmanska V, Quadrianto N, Lampert CH (2012) Augmented attribute representations. In: Computer VisionECCV 2012, vol 7576, pp 242–255. Springer
27. Shi R, Lee CH, Chua TS (2007) Enhancing image annotation by integrating concept ontology and text-based bayesian learning model. In: Proceedings of the 15th international conference on Multimedia, pp 341–344. ACM
28. Siddiquie B, Feris RS, Davis LS (2011) Image ranking and retrieval based on multi-attribute queries
29. Srikanth M, Varner J, Bowden M, Moldovan D (2005) Exploiting ontologies for automatic image annotation. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp 552–558. ACM
30. Srikanth M, Varner J, Bowden M, Moldovan D (2005) Exploiting ontologies for automatic image annotation. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp 552–558. ACM
31. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Ser B (Statistical Methodology) 73(3):273–282
32. Tousch AM, Herbin S, Audibert JY (2012) Semantic hierarchies for image annotation: A survey. Pattern Recogn Lett 45(1):333–345
33. Wang C, Yan S, Zhang HJ (2009) Large scale natural image classification by sparsity exploration. In: Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, pp 3709–3712. IEEE
34. Wang M, Yang K, Hua XS, Zhang HJ (2010) Towards a relevant and diverse search of social images. IEEE Trans Multimedia 12(8):829–842
35. Wang Y, Mori G (2010) A discriminative latent model of object classes and attributes. In: Computer Vision ECCV 2010, vol 6315, pp 155–168. Springer
36. Yu FX, Cao L, Feris RS, Smith JR, Chang SF (2013) Designing category-level attributes for discriminative visual recognition. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp 771–778. IEEE

37. Yu FX, Ji R, Tsai MH, Ye G, Chang SF (2012) Weak attributes for large-scale image retrieval. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp 2949–2956. IEEE
38. Yu X, Aloimonos Y (2010) Attribute-based transfer learning for object categorization with zero/one training example. In: Computer VisionECCV 2010, pp 127–140. Springer
39. Zhang H, Zha ZJ, Yang Y, Yan S, Gao Y, Chua TS (2013) Attribute-augmented semantic hierarchy: towards bridging semantic gap and intention gap in image retrieval. In: Proceedings of the 21st ACM international conference on Multimedia, pp 33–42. ACM



Chao Li is a graduate student at the School of Computer Science and Technology, Tianjin University, Tianjin, China. He received her B.S. degree from Tianjin University, Tianjin, China. His research interests include computer vision, multimedia retrieval.

Zhiyong Feng received the Ph.D. degree in mechanical Engineering from Tianjin University in 1996. He is currently a Professor of Computer Science and Associate Dean for School of Computer Science and Technology, Tianjin University. His research is in the field of software engineering, knowledge Engineering and distributed software. He is the senior member of China Computer Federation (CCF), and serves as supervisor of Supervisory Committee, the executive member of Education committee and the member of Software Engineering Committee of CCF. He is the member of Association for Computing Machinery (ACM) and the member of Institute of Electrical and Electronics Engineers(IEEE). He also serves as the executive member of Computer Higher Education Research Association of China.

Yahong Han received the Ph.D. degree from Zhe-jiang University, Hangzhou, China. He is currently an Associate Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. His current research interests include multimedia analysis, retrieval, and machine learning.