Towards Cross-Category Knowledge Propagation for Learning Visual Concepts

Guo-Jun Qi[†], Charu Aggarwal[♯], Yong Rui[‡], Qi Tian[♮], Shiyu Chang[†] and Thomas Huang[†]
Beckman Institute, University of Illinois at Urbana-Champaign[†]
IBM T.J. Watson Research Center[♯], Microsoft Advanced Technology Center[‡]
Department of Computer Science, University of Texas at San Antonio[♮]
{qi4,chang87,t-huang1}@illinois.edu[†], charu@us.ibm.com[♯]
yongrui@microsoft.com[‡], qitian@cs.utsa.edu[♮]

Abstract

In recent years, knowledge transfer algorithms have become one of most the active research areas in learning visual concepts. Most of the existing learning algorithms focuses on leveraging the knowledge transfer process which is specific to a given category. However, in many cases, such a process may not be very effective when a particular target category has very few samples. In such cases, it is interesting to examine, whether it is feasible to use **cross-category knowledge** *for improving the learning process by exploring the knowledge in correlated categories. Such a task can be quite challenging due to variations in semantic similarities and differences between categories, which could either help or hinder the cross-category learning process. In order to address this challenge, we develop a cross-category label propagation algorithm, which can directly propagate the inter-category knowledge* **at instance level** *between the source and the target categories. Furthermore, this algorithm can automatically detect conditions under which the transfer process can be detrimental to the learning process. This provides us a way to know when the transfer of cross-category knowledge is both useful and desirable. We present experimental results on real image and video data sets in order to demonstrate the effectiveness of our approach.*

1. Introduction

While image and video categorization algorithms have seen tremendous advancements in recent years, the small number of training examples continues to remain a challenge for the learning algorithms, especially in the context of complex semantic concepts. One possible solution is to extract additional knowledge from other sources which describe the same set of categories using either different feature vectors or a different data set. This process is known as transfer learning [2] [14] [11] [8], and has been applied to extract the knowledge from related domains to enhance the learning process. Such methods typically make use of knowledge between different domains or data sets which relate to the *same concept category*. For example, one may use information in a source data set (e.g., Caltech 101) to model the same category in the target data set (e.g., Flickr web images) [13][2], or transfer the information from one modality (e.g., text modality) to the target modality (e.g., image modality) [6]. These algorithms mainly focus on leveraging the knowledge *within* the category for domain adaptation. This approach becomes less effective *when a particular target category has limited examples across different data sources, and/or is too complex to be individually modeled from the category of itself.*

To overcome this difficulty, in this paper, we develop an approach for cross-category transfer learning for a visual classification task. The basic assumption for the cross-category transfer learning process (*CCTL*) is that once we have a large number of source categories, the modeling of the target category can become much easier with the extra information from other categories [11]. A key observation is that semantic concepts do not exist independently, because most of them are closely correlated with each other. This makes it possible to transfer cross-category knowledge. It also means that it is sometimes possible to learn a brand new concept with limited supervision information, when we have a large pool of categories with wide semantic coverage. Such cross-category knowledge not only exists in positive correlated categories, but also exits in negatively correlated categories. Moreover, when modeling a complex concept with large intra-class variants, it is often difficult to model it directly without sufficient training examples. In this case, we develop a divide-and-conquer framework to overcome the difficulty effectively.

In order to design such a cross-category transfer learning algorithm, it is critical to know how and when to perform cross-category transfer learning. Our approach is unique in

its design to maximize the beneficial cross-category transfers during the learning process.

How to do cross-category transfer. Most of the available algorithms for transfer learning are designed in the context of *domain adaptation* [13], in which a set of pre-learned models are used as a prior to adapt into a target domain with fewer examples. For example, the work in [13][11] are both designed for cross-domain adaptation by constraining the classification hyperplane in the target domain to be close to that in the source domain. The work in [14] proposes a two phase transfer approach by identifying which models from various sources can be reused to improve the target classifier. While these algorithms work well in the domain adaption scenario, they do not reveal the intrinsic category correlations among different categories. This can be detrimental to cross-category knowledge propagation, since blind knowledge transfer can actually hurt the learning process. To overcome this problem, we propose to explicitly learn and leverage cross-category label correlations in order to transfer knowledge from different source categories to the target category.

When to do cross-category transfer. Transfer learning can sometimes have detrimental effects [9], when the knowledge propagation is noisy. In the context of the cross-category transfer problem, this would correspond to a scenario in which some source categories do not contain helpful transfer knowledge, which when used inappropriately, can harm the modeling of the target category. To avoid such negative transfer, we follow the *lazy* principle of *never launching the cross-category transfer process unless it is necessary.* A data-driven method is proposed to automatically select the best category for transfer which minimizes the learning error. When no source categories have relevant information, the transfer process is not performed, and a non-transfer model is launched to avoid negative transfer.

Next we give a concise overall of the most related work.

1.1. Related Work

Transfer learning methods can be categorized into several types of approaches. Some of the earliest transfer learning algorithms concentrate on transferring knowledge to narrow distributive difference of training and testing data [2][10]. Source instances are directly used to train the weak hypotheses by combining the training and testing samples in each iteration [13][14]. Another type of transfer learning algorithm aims at transferring the knowledge between heterogeneous domains [6]. For example, the method in [6] designs a transfer learning process to propagate knowledge between the text and image domain via cross-domain translators. Comprehensive reviews of the transfer learning algorithms can be found in [5].

To the best of our knowledge, there are no known techniques for *directly* performing cross-category learning, which is the focus of this paper. There is however one possible indirect approach by combining multiple source classifiers [14], where the source models are pre-learned in a separate phase. However, these algorithms combine the source categories without explicitly modeling the intrinsic label correlations between different categories. This is one of main reasons for negative transfer since not all the source categories contain valuable (instead of harmful) label correlations. On the contrary, we aim to construct the cross-category classifier with label correlations. The transfer process is explicitly enforced to align with the label correlation between different categories.

It is worth noting that the previous work in correlative multi-label (*CML*) classifiers [7] model the label correlation. However, as stated in [7], not all the label correlations between categories are helpful for classification, and some even contain negative information that is detrimental to multi-label classification. Furthermore, *CML* is limited in ignoring the category correlations among different samples. It only explores the category correlation within the same sample. The work in [8] proposes to extract the concept relatedness from the text knowledge base. However, the text domain knowledge can differ from that in the visual domain and the concept relatedness in text domain can be distorted compared to that in the visual domain. The *CCTL* overcomes the above problems, where the transfer function can be constructed by directly mining the label correlation across different samples in visual knowledge base. This greatly increases its flexibility and ability to transfer knowledge across different categories. We will show the advantage of this approach over existing methods in the experimental section.

To summarize, to our best knowledge, this paper proposes the first *direct* cross-category transfer learning method which greatly improve the small sample learning process. We will demonstrate the superiority of this approach over existing methods in Sections 2 and 3 via theoretical analysis and in Section 4 via extensive experiments and comparisons.

2. Cross-Category Classification Process

We denote the source sets by $\mathcal{A}_l = \{(\mathbf{x}_{l,n}, y_{l,n}) \,|n = 1, \cdots, N_l\}$ for $l = 1, \cdots, L$ over L source categories. The variables $\mathbf{x}_{l,n}$ and $y_{l,n}$ represent the feature vector and the ground truth label for the nth instance of the lth category, and N_l is the number of source examples for the lth category. For simplicity in exposition, we assume that the label $y_{l,n} \in \{+1, -1\}$ is binary, though the extension to the general case is straightforward. In addition, a (small) training set $\mathcal{T} = \{(\mathbf{x}_i, y_i) \,|i = 1, \cdots, N\}$ of the target category is available for learning purpose. It is worth noting that instances in the training sets $\mathcal{T}$ can be different from those in the source sets $\mathcal{A}_l$. In other

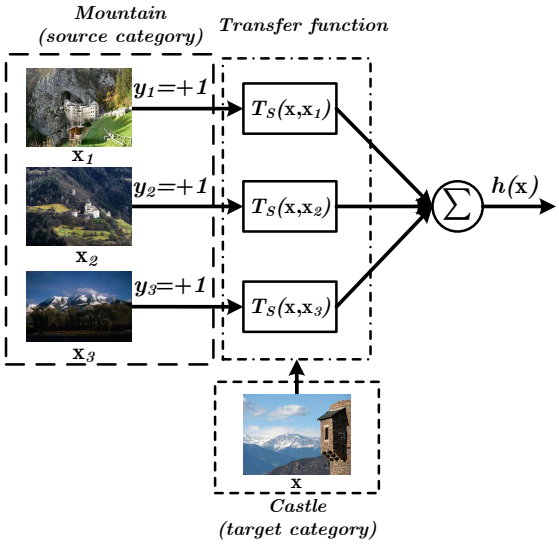


Figure 1. Illustration of cross-category label propagation as Equation (1). The labels in source category 'mountain' are propagated to target category 'castle.'

words, the set $\{\mathbf{x}_{l,n}|n = 1, \cdots, N_l, l = 1, \cdots, L\}$ is not necessarily the same as $\{\mathbf{x}_i|i = 1, \cdots, N\}$. This increases the flexibility and usability of the approach for different scenarios.

The key ingredient in the proposed approach is a classifier capable of transferring the cross-category labeling information. For this purpose, we define a real-valued transfer function $T_S(\mathbf{x}, \mathbf{x}_{l,n})$ to connect the lth source category and the target category. Then we propose a cross-category label propagation approach (CCLP) in order to learn the cross-category classifier, in terms of this transfer function. Specifically, this classifier propagates the labels from the instances in $\mathcal{A}_l$ to the target category to form a discriminant function $h_l(\mathbf{x})$, whose sign indicates the label of the sample $\mathbf{x}$ in the target category. The relationship of the discriminant function with the transfer function is as follows:

$$h_l(\mathbf{x}) = \frac{1}{|\mathcal{A}_l|} \sum_{\mathbf{x}_{l,n} \in \mathcal{A}_l} y_{l,n} T_S(\mathbf{x}, \mathbf{x}_{l,n}) \tag{1}$$

Here, $|\mathcal{A}_l|$ is the cardinality of $\mathcal{A}_l$.

Next, we discuss how the transfer function is defined. In many cases, the binary labels on the two categories can be different even on the same image, depending upon how they are collected or labeled. Hence, we use a function $\phi_S(\mathbf{x}, \mathbf{x}_{l,n})$ to measure the correlations between two different categories. In this paper we adopt $\phi_S(\mathbf{x}, \mathbf{x}_{l,n}) = \mathbf{x}^T S \mathbf{x}_{l,n}$ with the parametric matrix S. The choice of a proper matrix S is critical to the learning process. The label correlation function can be either positive or negative, representing either positive or negative correlations between categories. Note that the category correlation is a function of samples, so that the correlations on different samples can be different. In addition, we define the kernel function $k(\mathbf{x}, \mathbf{x}_{l,n})$ to measure the sample similarity. The source examples similar to the target one are more heavily weighted in CCLP process than the dissimilar ones, as the similar examples ought to contain more valuable information on the target example for knowledge propagation. The label transfer function is defined in terms of these two functions:

$$T_S(\mathbf{x}, \mathbf{x}_{l,n}) = \phi_S(\mathbf{x}, \mathbf{x}_{l,n})\, k(\mathbf{x}, \mathbf{x}_{l,n}) \tag{2}$$

Conventional label propagation algorithms propagate the labels within the same category, and the propagation weights are determined beforehand by the similarities between samples. On the other hand, the CCLP approach uses not only the sample similarities, but also the label correlations across different categories. This makes CCLP more effective in propagating the knowledge across different samples and categories. Figure 1 illustrates the idea of cross-category label propagation, and shows how the labels are propagated between 'mountain' and 'castle' between different images.

Here, we learn a weighted cross-category classifier with $\mathbf{w}_i$ as the sample weight for the i-th sample. Then the parameter S for each discriminant function h_l should be learned by minimizing the following objective function:

$$S^\star = \arg\min_S \Omega_l(S) \tag{3}$$

where

$$\begin{aligned}\Omega_l(S) &= \sum_{i=1}^{N} \mathbf{w}_i \left(1 - y_i h_l(\mathbf{x}_i)\right)_+ + \frac{\lambda}{2}||S||_F^2 \\ &= \sum_{i=1}^{N} \mathbf{w}_i \left(1 - \sum_{\mathbf{x}_{l,n} \in A_l} \frac{y_i y_{l,n} T_S(\mathbf{x}_i, \mathbf{x}_{l,n})}{|A_l|}\right)_+ + \frac{\lambda}{2}||S||_F^2\end{aligned} \tag{4}$$

where $(\cdot)_+ = \max(0, x)$, the $\|S\|_F$ is the Frobenius norm of the matrix S which serves as the regularization term, and λ is the balancing parameter. The terms in the objective function $\Omega_l(S)$ require some further explanation.

- The first term is the empirical loss of predictions made by the discriminant function h_l on the training set $\mathcal{T}$. This empirical loss is weighted by the sample weights $\mathbf{w}_i$ on the training set $\mathcal{T}$. We adopt the hinge loss to measure the cost. Based on the large margin principle, the hinge loss is minimized by maximizing the margin $y_i h_l(\mathbf{x}_i)$ of each training instance.
- The term $y_i y_{l,n} T_S(\mathbf{x}_i, \mathbf{x}_{l,n})$ in $(\cdot)_+$ enforces that the transfer function $T_S(\mathbf{x}_i, \mathbf{x}_{l,n})$ is aligned with the label correlation $y_i y_{l,n}$. In other words, when the source and target labels have positive correlation (i.e., $y_i y_{l,n} = +1$), the transfer function will be as positive as possible and vice-versa. This is a mechanism for capturing the correlations in the transfer function.

Before advancing into the optimization detail for the above objective function, we would like to briefly compare two kinds of across-category label correlations.

Category-Level Label Correlation. Most of existing related work, such as multi-label learning [7][12], models the inter-category correlations on the category level. It assumes that label correlations are irrelevant to the instances under consideration. However, in many cases the label correlations can vary across different instances. For example, 'mountain' and 'castle' can either co-occur with positive correlations when the castle is built by the mountain or be mutually exclusive with negative correlations in some other cases. In this case, assuming the same label correlation for all images can fail to capture such subtle inter-category correlation.

Instance-Level Label Correlation. On the contrary, the proposed model assumes the varying inter-category correlations across instances. Specifically, it measures cross-category relation with the function ϕ_S of a pair of source and target instances. For this purpose, intuitively the instance-level label correlation changes smoothly with slight changes of input instances. ϕ_S satisfies this requirement by imposing the Frobenius regularization on S for a smooth correlation function. It is worth noting that the smoothness requirement excludes the possibility of assigning an individual label correlation to each source instance (i.e., each $\phi_S(\mathbf{x}, \mathbf{x}_{l,n})$ becomes a constant coefficient $\phi_{l,n}$) since the neighboring source instances could have arbitrarily different label correlations in this case.

Now we move to the detail of optimizing the objective function (3). The use of a complete matrix S (with a large number of parameters) to parameterize the transfer function may lead to overfitting. Therefore, we restrict the matrix S to be diagonal in order to reduce overfitting. In other words, we have $S = \mathrm{diag}(\boldsymbol{\eta})$ with its diagonal elements in the vector $\boldsymbol{\eta}$. By formulating the Lagrangian function, we can obtain the following dual problem:

$$\boldsymbol{\beta}^\star = \underset{0 \le \beta_i \le \mathbf{w}_i}{\arg\max}\, \Xi_l(\boldsymbol{\beta}) \tag{5}$$

The corresponding dual objective is as follows:

$$\Xi_l(\boldsymbol{\beta}) = \boldsymbol{\beta}^T \mathbf{e} - \frac{1}{2\lambda} \boldsymbol{\beta}^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} \boldsymbol{\beta} \tag{6}$$

where the variables $\boldsymbol{\beta}$ and $\mathbf{e}$ both represent $N \times 1$ vectors, in which $\boldsymbol{\beta}$ contains the dual variables β_i as its elements and $\mathbf{e}$ is a vector containing unit values. Here $\boldsymbol{\Gamma}$ is a constant matrix with column vectors $\boldsymbol{\gamma}_i, 1 \le i \le N$ as follows:

$$\boldsymbol{\gamma}_i = y_i \sum_{\mathbf{x}_{l,n} \in A_l} \frac{y_{l,n} \mathbf{x}_i \circ \mathbf{x}_{l,n} k(\mathbf{x}_i, \mathbf{x}_{l,n})}{|\mathcal{A}_l|}; \tag{7}$$

where $\circ$ denotes the element-wise product of two vectors. Due to space constraints, we omit the derivation of this dual problem here.

The dual problem (5) can be solved by any quadratic programming solver, and with the optimal dual variables $\boldsymbol{\beta}^\star$ can be used to determine the corresponding primal solution:

$$\boldsymbol{\eta}^\star = \frac{1}{\lambda} \boldsymbol{\Gamma} \boldsymbol{\beta}^\star = \frac{1}{\lambda} \sum_{i=1}^{N} \beta_i^\star y_i \sum_{\mathbf{x}_{l,n} \in \mathcal{A}_l} \frac{y_{l,n} \mathbf{x}_i \circ \mathbf{x}_{l,n} k(\mathbf{x}_i, \mathbf{x}_{l,n})}{|\mathcal{A}_l|} \tag{8}$$

3. Learning the Cross-Category Ensemble

In the previous section, we address how to optimally transfer cross-category knowledge from a single category, which is the key contribution of this paper. In this section, we show that cross-category classifiers can be easily combined to integrate the knowledge from multiple source categories simply in an *Adaboost* [3] framework.

Algorithm 1 presents the learning procedure. Here $\mathrm{Learn\_CCC}(\mathcal{A}_l, \mathcal{T}, \mathbf{w})$ denotes the learning algorithm of cross-category classifier as described in the last section, with the source set $\mathcal{A}_l$, the training set $\mathcal{T}$ and weighting vector $\mathbf{w}$ as its input. In each iteration, L cross-category classifiers $h_l^{(t)}(\mathbf{x})$ are learned from each source category for $l = 1, \cdots, L$ as in Step 3. In addition to these L cross-category classifiers, an intra-category classifier $h_0^{(t)}$ is learned by propagating the labels in the target training set $\mathcal{T}$ to itself in Step 2. The function $h_0^{(t)}$ does not transfer any cross-category information, since it only uses the intra-category labels without any cross-category information, and thus it learns a non-transfer classifier.

Then, these $L+1$ candidate classifiers compete with each other, and the one with the least learning error is selected. The sample weighting vector $\mathbf{w}^{(t)}$ in each iteration usually specifies an aspect of the target category. Through the competition, the source category with the best correlation will win in the competition and is selected to model the target category in each iteration. We note that if the intra-category classifier $h_0^{(t)}$ wins, it indicates that no source category can better model the target category than itself in the current iteration. In such a case, $h_0^{(t)}$ plays a role of a *safety valve* in avoiding negative transfer by intelligently switching to a non-transfer approach where it is appropriate. Such an automatic determination of "when to transfer" is critical for constructing a robust classifier.

4. Experiments

In this section, we present experiments comparing the proposed *CCTL* approach with the baseline *AdaBoost* algorithm, existing transfer learning algorithms as well as a multi-label classifier. The experiments are designed to answer the following questions.

1. When and how are the source categories used for modeling the target category? We demonstrate the varia-

Algorithm 1 Learning cross-category ensemble

input source sets $\mathcal{A}_l, l = 0, 1, \cdots, L$, the training set $\mathcal{T}$ and the number of iterations T.

1 Initialize sample weights: $\mathbf{w}_i^{(0)} = 1/N$ for all $i = 1, 2, \cdots, N$, and $f^{(0)} = 0$.

for $t = 1, \cdots, T$ **do**

2 Train $h_0^{(t)} \leftarrow$ **Learn_CCC**$(\mathcal{T}, \mathcal{T}, \mathbf{w}^{(t)})$, and calculate the weighted learning error ε_0 according to $\mathbf{w}^{(t)}$.

for $l = 1, \cdots, L$ **do**

3 Train $h_l^{(t)} \leftarrow$ **Learn_CCC**$(\mathcal{A}_l, \mathcal{T}, \mathbf{w}^{(t)})$, and calculate the weighted learning error ε_l.

end for

4 Pick the model $h^{(t)} = h_j^{(t)}$ with the minimum training error, where $j = \underset{i=0,1,\cdots,L}{\arg\min}\ \varepsilon_i$.

5 Set $\alpha^{(t)} = \frac{1}{2} \log \frac{1 - \varepsilon_j}{\varepsilon_j}$.

6 Update , $\mathbf{w}_i^{(t+1)} = \mathbf{w}_i^{(t)} \exp\left\{-\alpha^{(t)} y_i \mathrm{sgn}(h^{(t)}(\mathbf{x}_i))\right\} \Big/ Z^{(t)}$ where $Z^{(t)}$ is a normalization constant such that $\sum_{i=1}^{N} \mathbf{w}_i^{(t+1)} = 1$.

7 Update $f^{(t)} = f^{(t-1)} + \alpha^{(t)} \mathrm{sgn}(h^{(t)}(\mathbf{x}))$.

end for

output the final classifier $\mathrm{sgn}(f^{(T)}(\mathbf{x}))$.

tions in performance with varying number of source categories and show how the source categories are combined to represent the target category.

2. How well does the *CCTL* method perform with an extremely small number of training examples? Therefore, we test the accuracy with varying number of training examples.

4.1. Data Sets

We compare the algorithms on two real data sets to evaluate their performances. The data sets are described below.

Flickr scene image data set. The first data set is a publicly-available natural scene image data set crawled from Flickr.com [1]. It contains 17,463 training images and another 17,463 testing images. Figure 2 illustrates some example images. There are 33 scene categories defined on this data set. Figure 3 shows the number of positive examples on these 33 categories. The 10 categories with the fewest positive examples are selected as the target categories, and the remaining 23 categories are used as the source categories. Each of these 10 target categories contain at most 245 positive examples. For each image, 500-dimensional bag-of-visual-words feature vectors are extracted. First, the Difference of Gaussian filter is on the gray scale images to detect a set of key-points and scales respectively; then the Scale Invariant Feature Transform (SIFT) is computed over the local region defined by the key-point and scale. The vector quantization on SIFT region descriptors is performed to construct the visual vocabulary by exploiting the k-means

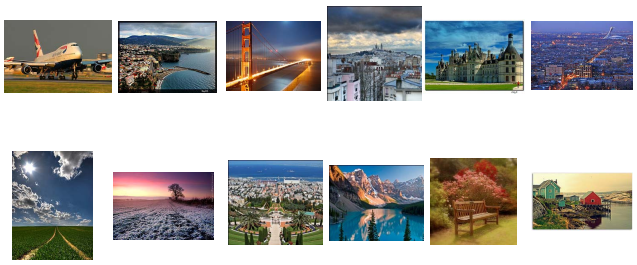


Figure 2. Example images in Flickr scene image data set.

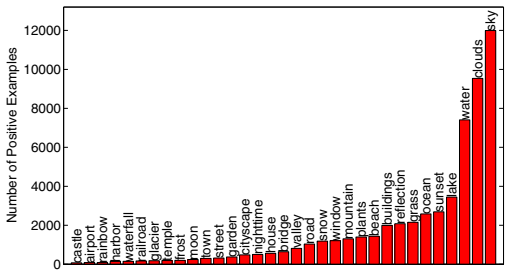


Figure 3. The numbers of positive examples over 33 categories in Flickr scene image data set. The fewest 10 categories are selected as the target categories while the remaining 23 categories are used as the source categories.

clustering that generates 500 visual words[1].

LSCOM video dataset. The other data set is a public video data set - LSCOM video dataset [4]. It contains 85 hours of international broadcast videos from Arabic, Chinese, and US news sources. The whole video corpus is automatically segmented into 61,901 portions, of which 43,331 segments are used for training and 18,570 for testing. 39 semantic categories are annotated on the dataset, which are related to program categories, setting, people, objects, activities, events, and graphics. Figure 4 shows the numbers of positive examples on these 39 categories on LSCOM video data set. The 10 categories with the fewest positive examples are selected as the target categories and the remaining 29 categories are used as source categories. On each key frame of the segmented shots, a 64-d color histogram, a 144-d color correlogram, a 73-d edge direction histogram, a 128-d wavelet texture and 225-d block-wise color moments are extracted and concatenated into a 634-dimensional feature vector per segment.

4.2. Experimental Setting

We compare our proposed *CCTL* approach with three baseline algorithms. (a) The *AdaBoost* algorithm [3] directly learns with the training set of the target category. (b) The *TaskTrAdaBoost* (Task Transfer *Adaboost*) algorithm[14] uses transfer learning to combine a set of pre-

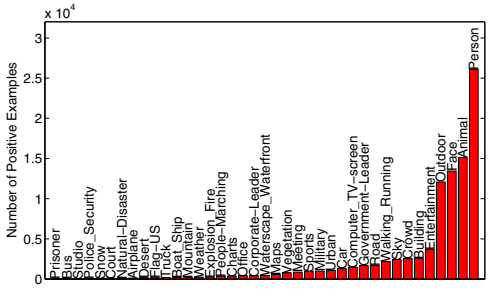


Figure 4. The numbers of positive examples over 39 categories in LSCOM video data set. The fewest 10 categories are selected as the target categories while the remaining 29 categories are used as the source categories.

learned classifiers. In the experiment, SVM classifiers are trained as the base classifiers in *TaskTrAdaBoost*. For a fair comparison with *CCTL*, we give additional benefits to the SVM classifiers, where they are trained on both the source and target categories in *TaskTrAdaBoost*. (c) *CML* (Correlative Multi-Label) classifier [7] is a multi-label classifier which also explores the label correlations based on structural SVM.

For the sake of fair comparison, in the proposed *CCTL*, only the training examples are used in the source sets associated with the corresponding source categories. It guarantees that no more information is used for learning in *CCTL* as compared with *AdaBoost*, *TaskTrAdaBoost* and *CML*.

On the Flickr image dataset, χ^2 kernel function is used in the *CML* as well as the cross-category classifiers in the *CCTL*, *Adaboost* and *TaskTrAdaboost* since it has been reported with competitive performance on bag-of-words features. On the LSCOM video dataset, Gaussian kernel is adopted for the extracted features. For these algorithms, the parameters are determined via a 5-fold cross-validation process on the training set. Since each image can contain more than one label, the training and prediction are made in the binary setting. The widely used Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is reported for comparison.

4.3. Results

Table 1 compares the categorization performance of the different algorithms in terms of AUC on Flickr image dataset and LSCOM video dataset. The results are obtained by using the whole training set of the target categories. The results show the competitive advantage of *CCTL* over the other algorithms.

In Figure 5, we demonstrate the selected source categories during the first ten iterations of *CCTL* and their associated combination coefficients $\alpha^{(t)}$. We make the following observations:

Table 1. Comparison of different categorization algorithms over 10 target categories in Flickr image data set and LSCOM video data set in terms of AUC. The best performance for each category is highlighted in bold.

(a) Flickr Image Dataset

Category	*Adaboost*	*CML*	*TaskTrAdaboost*	*CCTL*
castle	0.6214	0.6898	0.6826	**0.7210**
airport	0.5799	0.6921	0.6965	**0.7420**
rainbow	0.5782	0.5523	0.5684	**0.6381**
harbor	0.5876	0.6821	**0.6885**	0.6742
waterfall	0.5038	0.7986	0.8003	**0.8550**
railroad	0.5489	0.5317	0.5418	**0.6101**
glacier	0.6981	0.7543	0.7612	**0.8096**
temple	0.4293	0.4547	0.4717	**0.5494**
frost	0.6307	0.6978	0.7532	**0.7824**
moon	0.5388	0.6885	0.6731	**0.7768**

(b) LSCOM Video Dataset

Category	*Adaboost*	*CML*	*TaskTrAdaboost*	*CCTL*
Prisoner	0.5583	0.502	0.6003	**0.7991**
Bus	**0.537**	0.4669	0.4493	0.5024
Studio	0.7525	0.8068	0.8162	**0.9019**
Police_Security	0.6235	0.6223	0.6898	**0.7528**
Snow	0.6344	0.7035	0.7652	**0.8254**
Court	0.6297	0.6641	0.6638	**0.7168**
Natural-Disaster	0.6228	0.6037	0.731	**0.7864**
Airplane	0.4253	0.7229	0.7824	**0.8114**
Desert	0.4414	0.5762	0.6962	**0.75**
Flag-US	0.683	0.6944	0.7284	**0.7416**

Table 2. Comparison of *CCTL* with and without intra-category classifier over the target 10 categories on Flickr scene image set.

Category	Without intra-category classifier	With intra-category classifier
castle	0.7210	0.7210
airport	0.7253	**0.742**
rainbow	0.6381	0.6381
harbor	0.6714	**0.6742**
waterfall	0.7895	**0.855**
railroad	0.5983	**0.6101**
glacier	0.7781	**0.8096**
temple	0.5261	**0.5494**
frost	0.7800	**0.7824**
moon	0.7401	**0.7768**

1. The intra-category classifiers are usually selected in the first iteration to initialize the *CCTL* process. Beginning from the second iteration, the cross-category classifiers are usually selected. In most cases, it selects the source categories with positive correlations. For example, for the target category 'castle,' the source categories 'grass' and 'building' are selected in the second and fourth iteration, since 'castle' is a particular type of 'building' and often encompassed by 'grass.' In addition, the categories with negative correlations are used to transfer the labeling information. For example, 'valley' does not occur with 'airport,' but it is selected in

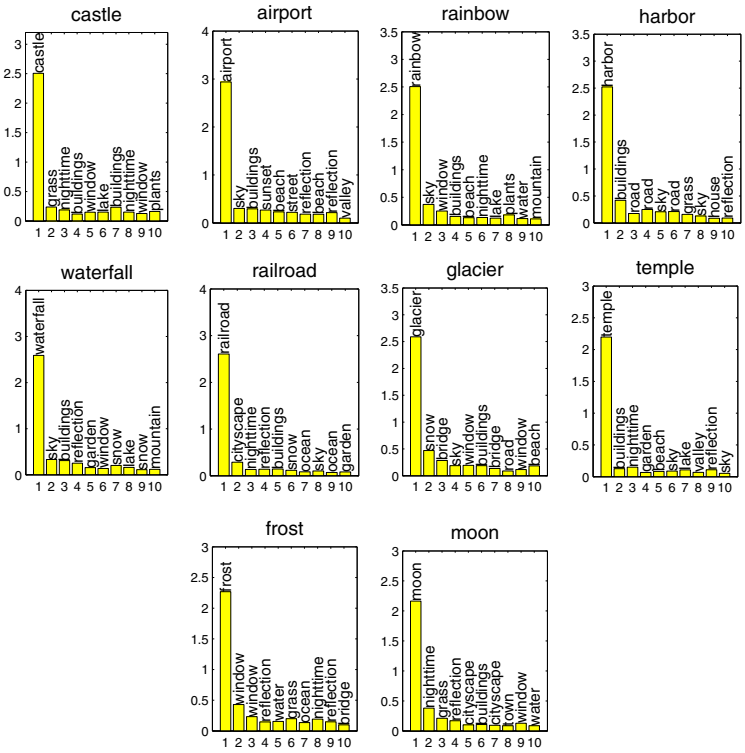


Figure 5. Illustration of selected source categories in the first 10 iterations on Flickr scene image set. In each subfigure, the horizontal axis is the iteration number, and the vertical axis depicts the weights $\alpha^{(t)}$ for each classifier. Above the bars are the names of selected categories for transfer in each iteration.

the tenth iteration when modeling 'airport.'

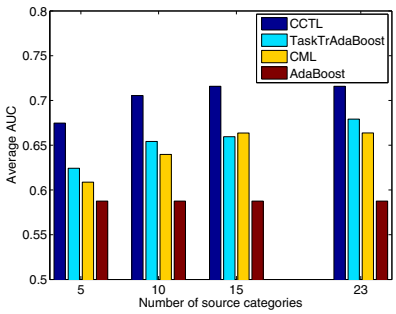
2. The coefficients $\alpha^{(t)}$ of the intra-category classifier is larger than those of the successive cross-category classifiers. However, this does not mean that cross-category classifiers are not important. The trained classifiers in the first iteration often focus on classifying the negative examples which are dominating over the training set. However, in real applications, the true positive predictions are often more important, and successive cross-category classifiers gradually concentrate on classifying the positive examples to increase the true positive rates.

To show the effect of the intra-category classifier, we also conduct a comparison experiment for *CCTL* with and without intra-category classifiers. The results are presented in Table 2. It is evident that the *CCTL* with the intra-category classifier performs better than the *CCTL* without intra-category classifier. When the intra-category classifier is used, the performance can be improved by avoiding the negative transfer.

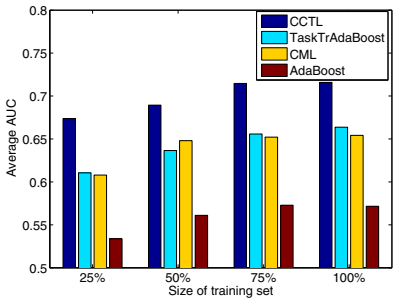
We also conduct a comparison with varying number of source categories and positive training examples. In Figure 6 (a), we compare the average AUC over all the ten target categories with varying number of source categories when all the training examples are used. The source categories are ordered by the number of positive examples here. It is evident that the categorization performances are improved from 0.67 with five source categories to 0.72 with the all 23 source categories. In Figure 6(b), we compare the average AUC with varying number of training examples of the target categories when all the 23 source categories are used. With much fewer training examples, we can observe that the *CCTL* continues to be much more robust than other algorithms by the help of the cross-category knowledge.

5. Conclusion

In this paper, we present an efficient method for cross-category knowledge transfer in the image and video domain, when there are only a small number of positive examples. This method is effective because it takes into ac-



(a) Comparison with varying numbers of source categories. All training examples are used in this comparison.



(b) Comparison with varying sizes of training set. All 23 source categories are used in this comparison.

Figure 6. Average AUC value over 10 target categories in Flickr scene image dataset with varying number of source categories (a) and sizes of training set (b).

count two key factors in cross-category knowledge transfer learning: **(a)** we directly model the category correlations between the source and target categories; and **(b)** transfer only if it can help. Specifically, we formulate cross-category label propagation processes which directly model the category correlations between the source and target categories. These base classifiers compete with one another in each iteration, and the most correlated category is selected to model the target category. We conduct extensive experiments using real data sets to compare the proposed approach against state-of-the-art techniques, and show its advantages over existing approaches in classification accuracy and computational efficiency.

Acknowledgment

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. Nus-wide: A real-world web image database from national university of singapore. In *ACM International Conference on Image and Video Retrieval*, July. 901

[2] W. Dai, Q. Yang, G. Xue, and Y. Yu. Boosting for transfer learning. In *Proceedings of International Conference on Machine Learning*, 2007. 897, 898

[3] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997. 900, 901

[4] M. R. Naphade, L. Kennedy, J. R. Kender, S.-F. Chang, J. R. Smith, P. Over, and A. Hauptmann. A light scale concept ontology for multimedia understanding for trecvid 2005. Technical report, IBM Research Technical Report, 2005. 901

[5] S. Pan and Q. Yang. A survey on transfer learning. Technical report, Hong Kong University of Science and Technology, Hong Kong, China, November 2008. 898

[6] G.-J. Qi, C. Aggarwal, and T. Huang. Towards cross-domain knowledge propagation from text corpus to web images. In *Proc. of International World Wide Web conference*, Hyderabad, India, March 28-April 1, 2011. 897, 898

[7] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang. Correlative multi-label video annotation. In *Proc. of International ACM Conference on Multimedia*, Augsburg, Germany, September 2007. 898, 900, 902

[8] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, and B. Schiele. What helps where - and why? semantic relatedness for knowledge transfer. In *Proceedings of Computer Vision and Pattern Recognition*, 2010. 897, 898

[9] M. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich. To transfer or not to transfer. In *NIPS 2005 Workshop on Transfer Learning*, 2005. 898

[10] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *Proceedings of ECCV*, 2010. 898

[11] T. Tommasi, F. Orabona, and B. Caputo. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In *Proceedings of Computer Vision and Pattern Recognition*, 2010. 897, 898

[12] R. Yan, J. Tesic, and J. R. Smith. Model-shared subspace boosting for multi-label classification. In *ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, 2007. 900

[13] J. Yang, R. Yan, and A. Hauptmann. Cross-domain video concept detection using adaptive svms. In *ACM Conference on Multimedia*, 2007. 897, 898

[14] Y. Yao and G. Doretto. Boosting for transfer learning with multiple sources. In *Proceedings of Computer Vision and Pattern Recognition*, 2010. 897, 898, 901