Learning Hypergraph-regularized Attribute Predictors

Sheng Huang[†], Mohamed Elhoseiny[‡] Ahmed Elgammal[‡] Dan Yang[†]
[†]College of Computer Science, Chongqing University, P.R.China. [‡]Department of Computer Science, Rutgers University, U.S.A.

Although a lot of impressive attribute learning approaches have been proposed in the recent decade, there still remain two problems unsolved in attribute learning [1]. The first issue is the consideration of the correlations among attributes and the second one is the exploitation of side information during attribute learning. In this paper, we present a novel attribute learning framework named Hypergraph-based Attribute Predictor (HAP) for addressing both of these two issues.

In HAP, a hypergraph is leveraged to depict the attribute relations among samples in which each hyperedge represents an attribute relation. Then, the attribute prediction problem can be translated as a supervised regularized hypergraph cut issue where its hypergraph cuts are deemed as the attribute predictions of samples. Finally, the attribute predictors can be jointly obtained by learning the mappings from the sample space to the attribute prediction space (the space of the learned hypergraph cuts). The processes details of HAP is summarized in Fig 1.

HAP Model: In HAP, a hypergraph $G(V,E)$ is defined for depicting the attribute relations of samples where V and E are the vertex set and hyperedge set respectively. In this hypergraph, the vertex $v_i \in V$ is corresponding to the sample $x_i \in X$ and each hyperedge is defined as a vertex set that shares the same attribute label. Then, we translate the attribute prediction issue as a regularized hypergraph cut issueand consider a collection of hypergraph cuts F as the attribute predictions of samples [4]. In such way, the optimal hypergraph cuts should not only preserve the structures of hyperedges (attribute relations) but also minimize the prediction errors of the train samples during hypergraph partition,

$$\begin{aligned}\hat{F} &= \arg\min_F \{\Omega(F,G) + \lambda\Delta(F,Y)\} \\ &= \mathrm{Tr}(F^T L_H F) + \lambda ||F-Y||^2, \end{aligned} \tag{1}$$

where $\Omega(F,G)$ is the attribute relation loss function, $\Delta(F,Y)$ is the attribute prediction loss function and λ is a positive parameter to reconcile these two losses. In this formula, L_H is the Laplacian matrix of the defined hypergraph G and Y is the collection of attribute labels.

Now, the problem of seeking attribute predictors can be transformed as a problem that finding a mapping B from the sample feature space to the attribute predictions space (constituted by the learned hypergraph cus), i.e. $F = X^T B$. We can introduce a L_2-norm constraint to B for avoiding the overfitting and then reformulate the Equation 1 as the following optimization problem to obtain the optimal B

$$\hat{B} = \arg\min_B (\mathrm{Tr}(B^T X L_H X^T B) + \lambda ||X^T B - Y||^2 + \eta ||B||^2). \tag{2}$$

where η is a positive regularization parameter.

CSHAP Model: By considering the regularized hypergraph cut issue as a multi-graph cut issue, we extend the HAP model to exploit the classification information during attribute learning,

$$\hat{B} = \arg\min_B (\mathrm{Tr}(B^T X L_W X^T B) + \lambda ||X^T B - Y||^2 + \eta ||B||^2) \tag{3}$$

where $L_W = L_H + \gamma L_*$ is the combination of the Laplacian matrices of the hypergraph, which encodes the attribute relations of samples, and a graph or hypergraph which encodes the class relations of samples. We name this new HAP model Class Specific HAP (CSHAP). The idea of CSHAP can be easily generalized for incorporating the other side information within HAP model just via accumulating more laplacian matrices of the hypergraphs or graphs which encode different side information.

Kernelization: The mapping from the feature space to the attribute prediction space may be not linear. This motivated us to present the kernelization

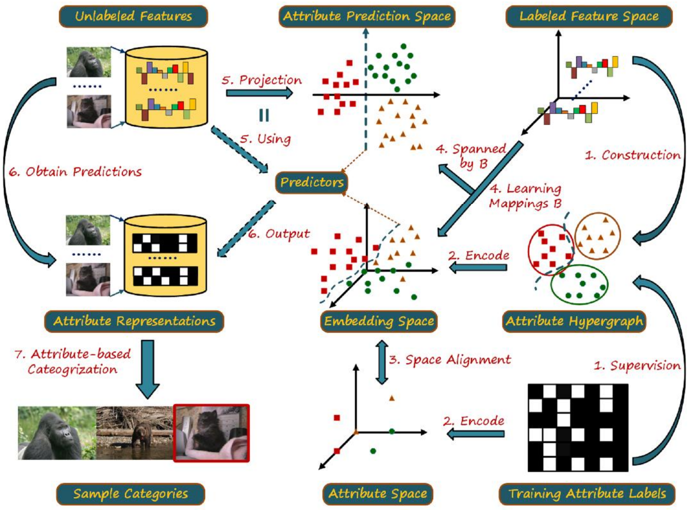

Figure 1: The overview of HAP.

for our method. According to the generalized representer theorem [3], the sample matrix X can be represented as a kernel matrix K_X via giving a kernel function $k(\cdot,\cdot)$. The kernelized attribute predictors B are the mappings from the kernelized feature space to attribute prediction space, i.e. $F = K_X^T B$. In such case, the objective functions of the Kernelized HAP (KHAP) and K-ernerlized CSHAP (KCSHAP) can be denoted as follows

$$\hat{B} = \arg\min_B \{\mathrm{Tr}(B^T K_X L_A K_X B) + \lambda ||K_X B - Y||^2 + \eta ||B||^2\} \tag{4}$$

where L_A is equal to L_H in the KHAP case and L_W in the KCSHAP case.

The models in Equation 2, 3 and 4 all can be easily solved as a regularized least square issue. At test time, given a unlabeled sample z_i, its attribute predictions can be achieved by projecting the sample into the subspace spanned by B, $\mathbf{p}_i = sign(z_i^T B)$ in linear case or $\mathbf{p}_i = sign\left(\sum_{j=1}^N k(z_i,x_j)^T B\right)$ in kernelized case, where $sign(\cdot)$ returns the sign of each element of a vector and $\mathbf{p}_i$ is a row vector encoded the predicted attributes of i-th sample.

Experimental Results: We experimented with three datasets: AWA, CUB and USAA. The results on attribute prediction, Zero/N-shot Learning, and categorization consistently validate the effectiveness of the proposed framework. For example, we record the attribute prediction accuracies of different attribute learning methods in Table 1. From these results, it is not hard to see that our approaches consistently outperform the compared methods.

The implementation details of our models and other experimental analysis can be referred to the full version of the paper at the Computer Vision Foundation webpage.

Table 1: Average Attribute Prediction Accuracies (in AUC).

Dataset	Prediction Accuracies (%)					
	HAP	**CSHAP**$_H$	**CSHAP**$_G$	DAP [2]	IAP [2]	ALE [1]
AWA	74.0	74.0	**74.3**	72.8/63.0*	72.1/73.8*	65.7
USAA	61.7±1.3	**62.2±0.8**	61.8±1.8	—	—	—
CUB	68.5	**68.7**	68.5	61.8	—	60.3

[1] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for attribute-based classification. In *CVPR*, pages 819–826, 2013.

[2] CH Lampert, H Nickisch, and S Harmeling. Attribute-based classification for zero-shot visual object categorization. *TPAMI*, 36(3):453, 2014.

[3] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. In *COLT*, 2001.

[4] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *NIPS*, pages 1601–1608, 2006.