METRIC LEARNING BASED ON ATTRIBUTE HYPERGRAPH

Yuchun Fang, Yandan Zheng*

School of Computer Engineering and Science, Shanghai University

ABSTRACT

In this paper, we propose an improved attribute hypergraph learning framework and adapt it for metric learning. Under the attribute hypergraph, each image is abstracted as a vertex and is contained in some hyperedges, each of which represents an attribute. The learned attribute hypergraph mines the correlation among multiple facial attributes and serves to reform the topology of the image similarity relationship in a database defined by the distance metric. By combining the merits of attribute and hypergraph, the reformed distance metric from the attribute hypergraph learning framework is able to capture the intrinsic information of the data set. Extensive experimental results on LFW data set confirm the effectiveness of the improved attribute hypergraph learning framework.

***Index Terms*—** Hypergraph learning, Metric learning, Human face retrieval, Semantic attribute

1. INTRODUCTION

Due to the popularity of photo sharing websites, such as Lofter and Facebook, a large amount of human facial images are available in our life. The rapid increase in the amount of face images has created a large number of requirements for many real-world applications, such as personalized image search, hair-style design, etc. For such large-scale face databases, how to measure the similarity of images has been a challenging problem. In traditional methods, the complex relationships among images are usually measured with distance metrics that can only evaluate pairwise relationships. Using these simple distance metrics, it is hard for high-level applications to make full use of the information hidden in the complex correlations net of images. Metric learning serves to refine and reform the general distance metric by revealing the profound relationships in an image database.

Metric learning is an attractive topic in the research area of image and multimedia retrieval. Pourdamghani et al. presented a new metric learning method using unlabeled data to estimate a nearest neighbor graph [1]. An improved penalized K-Nearest Neighbor graph metric was proposed in [2] based on the minimum of the average silhouette. These graph-based distance metrics effectively represent the topological structure residing in low intrinsic manifolds of high-dimensional image data. In the case of face images, multiple facial attributes can form very complex conceptional relations among images beyond pairwise ones. Hence, it is not sufficient to utilize a simple graph to represent the complex relations among face images.

Upon the foundation of graph theory and set theory, hypergraph is widely applied to image retrieval. Hypergraph has proven to be a good structure to encode complicated correlations [5, 6]. A hypergraph learning framework was proposed for image re-ranking [3]. In [4], a probabilistic hypergraph was used to represent the group relationships among images to build a content-based image retrieval system. Gao et al. [6] used multiple hypergraphs to capture the relationships of 3-D objects to avoid estimation of the distance among the 3-D objects at different granularities. Because of the good capability of capturing high order relationships, hypergraph is used for metric learning in this paper. The lower half of metric reforming block of Fig.1 illustrates how to construct a facial attribute hypergraph. The curve of each ellipse is a hyperedge that represents a facial attribute. Images at each hyperedge share the hyperedge attribute.

Attributes, rather than low-level features, are used to construct the hypergraph in this article. Due to the large visual variance of face image semantics, low-level features cannot be associated with high-level semantics, resulting in the so-called semantic gap. A promising route to narrow the semantic gap is exploiting a set of semantic descriptors. Attributes are adapted in our approach. As a type of intermediate-level feature representation of images, attributes endow more semantic meaning than low-level features. Attributes have received significant attention in recent years. Many studies have produced promising results using attributes such as object classification [7], image reranking [18], image retrieval [8, 19] and face verification [9].

Motivated by the above observations, we propose an improved hypergraph learning framework for metric learning. Fig.1 sketches the structure of our approach. We perform general low-level feature extraction to obtain an initial image representation in the first stage. In low-level feature spaces, machine learning based attribute models are adopted to extract semantic features. The hypergraph is constructed in the semantic spaces of image database. Hypergraph learning is designed to reform the topology of distance metric of the im-

*Corresponding author: ycfang@shu.edu.cn.

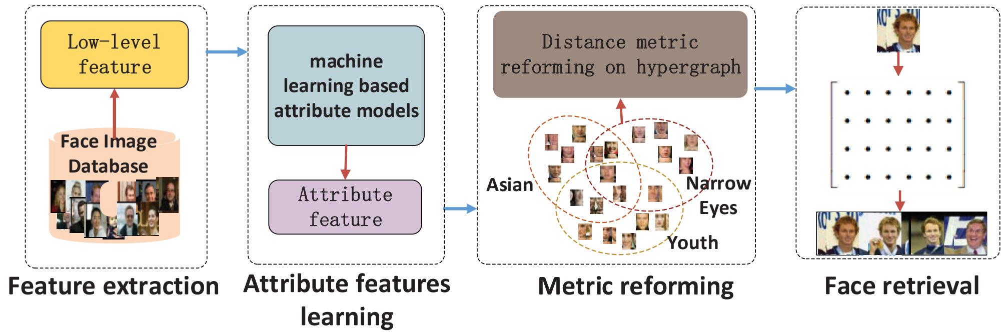

Fig. 1. The flowchart of our approach

age database and form a learned distance metric. The experiments results with the learned distance metrics validate the effectiveness of our approach.

2. RELATED WORK

In this section, we will briefly introduce the process of extracting facial attribute features and the theory of hypergraph.

2.1. Attribute Feature Representation

Face attributes are poses that are observable in images according to facial features. The attributes used in this paper are a portion of the attributes public in [9]. The selected 33 attributes, such as Oval Face, Chubby, are closely related to personal identification and appear in the facial region. However, attributes like Soft Lighting and Environment, which have no facial information, are discarded.

The attribute feature learning process for each image in the dataset contains two stages. The first stage is extracting the Uniform LBP (ULBP) [10] feature which is a very effective low-level feature in face recognition for each image. Low-level feature usually has high dimension which greatly influences the retrieval speed. As low-level feature contains high level semantic information, we use it to learn attribute feature. In the second stage, an attribute feature vector is learned from the attribute models using the ULBP feature. The machine learning based attribute models for each attribute are training on general support vector machine with SVM values provided in [9]. Using low-level feature, attribute feature is learned through machine learning based attribute models. The learned attribute feature vector contains 33 dimensions, each of which represents one of the selected attributes.

2.2. Hypergraph Analysis Theory

Conventional graph-based representations endow objects with pairwise relationships. Simply squeezing the complex relationships into pairwise ones can lead to a loss of useful information. Due to the over-simplicity of the graph, hypergraph is used. The definition of hypergraph was originally proposed by Berge and Minieka [11]. Below, we describe the theoretical basis for hypergraph.

Let $G = (V, E, \omega)$ be a hypergraph composed of a finite vertex set V, a hyperedge set E and the weights of the hyperedges ω. Each hyperedge has a weight $\omega(e)$. The hyperedge G can be represented by a $|V| \times |E|$ incidence matrix H with each entry defined in Eq.1.

$$h(v_i, e_j) = \begin{cases} 1, if v_i \in e_j \\ 0, otherwise \end{cases} \tag{1}$$

Based on the incidence matrix H, the degrees of vertex and hyperedge can be defined as $d(v) = \sum_{e \in E} \omega(e) h(v, e)$ and $\delta(e) = \sum_{v \in V} h(v, e)$.

We use D_v and D_e to represent the diagonal matrices of the vertex degrees and the hyperedge degrees, respectively. Let W denote the diagonal matrix of the hyperedge weights. Hypergraph has been proven to be beneficial to various tasks, including clustering [12], classification [17], and reranking [3].

Zhou et al. [13] proposed the normalized Laplacian method for hypergraph learning. The framework can be formulated as:

$$\arg \min_{f} \Omega(f) + \lambda R_{emp}(f) \tag{2}$$

where f represents the relevance score to be learned,$\Omega(f)$ denotes a normalized cost function, $R_{emp}(f)$ is an empirical loss, and λ is a parameter used to balance the regularizer with the empirical loss. Vertices sharing many incidental hyperedges can produce similar relevance scores through minimizing $\Omega(f)$. The regularizer of the hypergraph is defined as:

$$\Omega(f) = \frac{1}{2} \sum_{e \in E} \sum_{u,v \in e} \frac{\omega(e) h(u,e) h(v,e)}{\delta(e)} \times \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)^2 \tag{3}$$

Let $\Theta = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}}$, and $\triangle = I - \Theta$. The regularizer can be rewritten as $\Omega(f) = f^T \triangle f$, where the hypergraph Laplacian $\triangle$ is a positive semi-definite matrix.

3. ATTRIBUTE-BASED HYPERGRAPH LEARNING

A facial attribute hypergraph is proposed in this section to reform the topology of image distance metric. Each face image is regarded as a vertex of the hypergraph $G = (V, E, \omega)$. For a face image set containing n images and m attributes, let $V = \{v_1, v_2, ..., v_n\}$ represent n vertices and $E = \{e_1, e_2, ..., e_m\}$ denote m hyperedges. $\omega(e_i)$ is the weight of the hyperedge e_i, where $\sum_{i=1}^{m} \omega(e_i) = 1$ and $\omega(e_i) \geq 0$. In contrast to the traditional retrieval normalized cost function $\Omega(f)$, the vectors y and f need to be transformed into image distance metric $Y = \{Y_{i,j}, 1 \leq i, j \leq n\}$ and $F = \{F_{i,j}, 1 \leq i, j \leq n\}$, respectively. $Y_{i,j}$ represents the distance between image i with image j and $\mathbf{y_k}$ is a column of Y. While $F_{i,j}$ represents the reformed distance image i with image j and $\mathbf{f_k}$ is a column of F. Both Y and F are $n \times n$ metrics. Y is a distance metric computed through feature matrix and F is the topology reformed metric of Y. Therefore, the normalized cost function $\Omega(F)$ can be formulated as:

$$\begin{aligned} &\Omega(F) \\ &= \frac{1}{2} \sum_{k=1}^{n} \sum_{e \in E} \sum_{u,v \in e} \frac{\omega(e) h(u,e) h(v,e)}{\delta(e)} \left(\frac{F_{uk}}{\sqrt{d(u)}} - \frac{F_{vk}}{\sqrt{d(v)}}\right)^2 \\ &= \sum_{k=1}^{n} \mathbf{f_k^T} \triangle \mathbf{f_k} \end{aligned} \tag{4}$$

where I is the identity matrix and F is the reformed image distance metric. According to the learned attribute features, we define the affinity matrix A_s as:

$$\begin{aligned} A_s(i,j) &= \alpha \cdot A_{attribute} + \beta \cdot A_{global} \\ &= \alpha \cdot exp(-\frac{D_{attribute}(i,j)}{\bar{D}_{attribute}}) + \beta \cdot exp(-\frac{D_{global}(i,j)}{\bar{D}_{global}}) \end{aligned} \tag{5}$$

where $D_{attribute}$ and D_{global} are the Euclidean distances between v_i and v_j of attribute feature and global feature, respectively. Each entry in $A_s(i,j)$ represents the similarity between v_i and v_j. A_s is a $|V| \times |V|$ matrix and $A_s(i,j) \in [0,1]$. The hyperedge weight $\omega(e_k)$ can be computed as follow:

$$\omega(e_k) = \sum_{v_i \in e_k} \sum_{v_j \in e_k} A_s(i,j) / \sum_{r=1}^{m} \sum_{v_i \in e_r} \sum_{v_j \in e_r} A_s(i,j) \tag{6}$$

The empirical loss function can be computed as follow:

$$R_{emp}(F) = ||F - Y||^2 = \sum_{k=1}^{n} ||\mathbf{f_k} - \mathbf{y_k}||^2 \tag{7}$$

Considering that not all the attributes are available to the hypergraph, the formulation $\Psi(\omega)$ is integrated to Eq.2 to make a selection on the hyperedges.

According to the characteristics of the attributes used in the hypergraph in our method, the regularization framework can be reformulated as:

$$\underset{F,\omega}{\operatorname{argmin}}\, \Omega(F,\omega) + \lambda R_{emp}(F) + \mu \Psi(\omega) \tag{8}$$

The regularization framework can be rewritten as

$$\begin{aligned} &\underset{F,\omega}{\operatorname{argmin}}\, \Omega(F,\omega) = \underset{F,\omega}{\operatorname{argmin}} \{ \sum_{k=1}^{n} \mathbf{f_k^T} (I - \Theta) \mathbf{f_k} \\ &+ \lambda \sum_{k=1}^{n} ||\mathbf{f_k} - \mathbf{y_k}||^2 + \mu \sum_{i=1}^{m} \omega^2(e_i) \} \\ &s.t. \sum_{i=1}^{m} \omega(e_i) = 1, \lambda > 0 \end{aligned} \tag{9}$$

Eq.9 can be solved efficiently through alternating optimization. After ω is fixed, the regularization framework is used to differentiate $\Phi(F)$ with F, which is

$$\frac{\partial \Phi(F)}{\partial F} = \triangle F + \lambda (F - Y) = 0 \tag{10}$$

After some simple algebraic steps, we obtain

$$F = (I + \frac{1}{\lambda} \triangle)^{-1} Y \tag{11}$$

Next, we fix F and minimize ω,

$$\begin{aligned} &\underset{\omega}{\operatorname{argmin}} \{ \sum_{k=1}^{n} \mathbf{f_k^T} (I - \Theta) \mathbf{f_k} + \lambda \sum_{k=1}^{n} ||\mathbf{f_k} - \mathbf{y_k}||^2 \\ &+ \mu \sum_{i=1}^{m} \omega^2(e_i) \} \\ &s.t. \sum_{i=1}^{m} \omega(e_i) = 1, \lambda > 0 \end{aligned} \tag{12}$$

Through derivation of Φ with respect to $\omega(e_i)$, we have

$$\frac{\partial \Phi(F)}{\partial \omega(e_i)} = - \sum_{k=1}^{n} \mathbf{f_k^T} D_v^{-\frac{1}{2}} H R D_e^{-1} H^T D_v^{-\frac{1}{2}} \mathbf{f_k} + 2 \mu \omega(e_i) \tag{13}$$

where R is an $m \times m$ matrix with all entries equal to zero, except for $R(i,i) = 1$. The result of Eq.13 is

$$\omega(e_i) = \frac{\sum_{k=1}^{n} \mathbf{f_k^T} D_v^{-\frac{1}{2}} H R D_e^{-1} H^T D_v^{-\frac{1}{2}} \mathbf{f_k}}{\sum_{k=1}^{n} \mathbf{f_k^T} D_v^{-\frac{1}{2}} H D_e^{-1} H^T D_v^{-\frac{1}{2}} \mathbf{f_k}} \tag{14}$$

The iterative algorithm of the proposed process is summarized in Algorithm 1. Each step of the iterative process decreases the objective function $\Phi(F, \omega)$, which has a lower bound 0 [6]. Therefore, the convergence of the iterative process is guaranteed.

Algorithm 1 Attribute Hypergraph Learning

Input: A distance metric computed through feature matrix $Y = \{Y_{i,j}, 1 \leq i, j \leq n\}$

Output: Reformed distance metric $F = \{F_{i,j}, 1 \leq i, j \leq n\}$

Initialize similarity matrix A, incidence matrix H, vertex degree matrix D_v, hyperedge degree matrix D_e and initial weight matrix W.
Construct the hypergraph Laplacian $\triangle = I - \Theta$.
repeat
    Compute F according to Eq.11
    Update W according to Eq.14
    Update $\Theta = D_v^{-\frac{1}{2}} HWD_e^{-1}H^T D_v^{-\frac{1}{2}}$.
until W converge
return Reformed distance metric F

4. EXPERIMENTS

The experiment is conducted on the public dataset LFW [14] to analyze the performance of improved framework. LFW dataset consists of 13233 face images from 5749 different subjects. Since only 1680 subjects have more than one image, we discard those only with profile images and select four images per person as our experimental dataset.

For each image in the experimental dataset, the top $T = 4$ similar images are selected to compute the retrieval precision:

$$precision = \sum_{i=1}^{N} t_i/(N * T) \tag{15}$$

where t_i is the number of the top T similar images that are the same as the i-th images and N is the number of images int the experimental dataset.

We need to determine parameter λ in the hypergraph framework, while the other parameters can be directly computed from experimental data. According to the original work [13], we vary λ from 0 to 1 in the experiments. To evaluation the performance of the proposed attribute hypergraph learning framework in reforming the topology of image distance metric, four distance measures, including L2, L1, SCD [15] and Chi2 [16], are used in the experiment.

Fig.2 presents the precision of original and reformed distance metrics. The figure reports the precision bywith varying parameter λ. while tuning the balance parameter λ from 0.1 to 1, the precision curves initally increase and then decrease or remain stable. All the reformed distance metrices achieve comparable or better performance than the original distance metric. Furthermore, a parameter λ of (0.7, 0.9) results in better precision for the four distance measures.

Beyond the comparison of original and reformed distance metrics, we also evaluate the precision of the four distance measures. Fig.2 shows that the L1 distance metric is the most

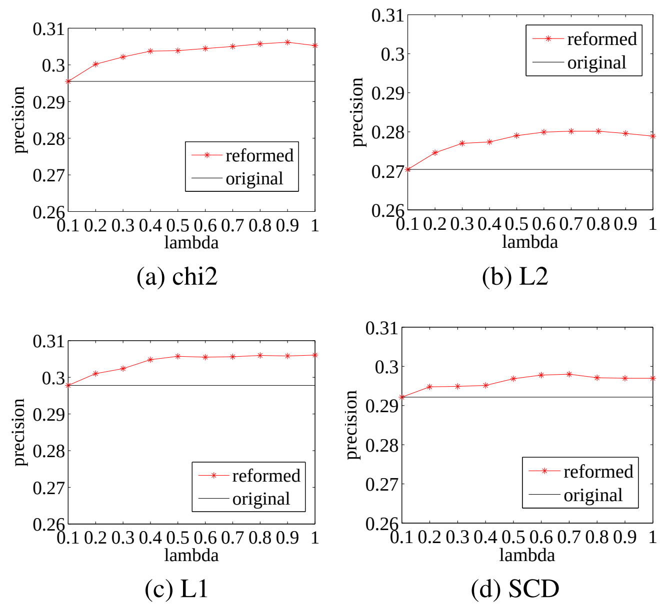

Fig. 2. The precision of reformed and original distance metric for the four distance measures.

suitable distance metric for attribute hypergraph based distance metric learning, since the L1 distance metric achieves the best precision among the distance metrics. The above experimental results reveal that attribute hypergraph is effective for reforming the topology of image similarity relationship in a database.

5. CONCLUSION

In this paper, an improved attribute hypergraph framework is proposed for distance metric learning. The improved method combines the strength of attribute and hypergraph in a unified framework. We build machine learning based attribute models to learn the attribute features. The learned attribute features enable us to model the relationships between images in experimental dataset using a hypergraph. Distance metric reforming is formulated as a learning tasks with the regularization framework on attribute hypergraph. The effectiveness of the reformed distance metrics on four distance measures are demonstrated through extensive experimentation.

6. ACKNOWLEDGMENTS

The work is funded by National Nature Science Foundation of China (No. 61170155) and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

7. REFERENCES

[1] N. Pourdamghani, H.R. Rabiee, and M. Zolfaghari. Metric learning for graph based semi-supervised human pose estimation[C]*International Conference on Pattern Recognition*. IEEE, 2012: 3386-3389.

[2] A.E. Baya and P.M. Granitto. Improved graph-based metrics for clustering high-dimensional datasets[C], *Ibero-American Conference on Artificial Intelligence*. Springer Berlin Heidelberg, 2010: 184-193.

[3] J. Cai, Z.J. Zha, M. Wang, and et al. An attribute-assisted reranking model for web image search[J],*IEEE Transactions on Image Processing*, 2015, 24(1): 261-272.

[4] Y. Huang, Q. Liu, S. Zhang, and et al. Image retrieval via probabilistic hypergraph ranking[C], *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on. IEEE, 2010: 3376-3383.

[5] S. Huang, M. Elhoseiny, A. Elgammal, and et al. Learning hypergraph-regularized attribute predictors[C], *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 409-417.

[6] Y. Gao, M. Wang, D. Tao, and et al. 3-d object retrieval and recognition with hypergraph analysis[J]. *IEEE Transactions on Image Processing*, 2012, 21(9): 4290-4303.

[7] W. Kusakunniran, S. Satoh, J. Zhang, and et al. Attribute-based learning for large scale object classification[C], *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2013: 1-6.

[8] B.C. Chen, Y.Y. Chen, Y.H. Kuo, and et al. Scalable face image retrieval using attribute-enhanced sparse codewords[J]. *IEEE Transactions on Multimedia*, 2013, 15(5): 1163-1173.

[9] N. Kumar, A.C. Berg, P.N. Belhumeur, et al. Attribute and simile classifiers for face verification[C], *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009: 365-372.

[10] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(12): 2037-2041.

[11] C. Berge. Graphs and hypergraphs[M],*Amsterdam: North-Holland publishing company*, 1973.

[12] S.R. Bulo and M. Pelillo. A game-theoretic approach to hypergraph clustering[C], *advances in neural information processing systems*. 2009: 1571-1579.

[13] D. Zhou, J. Huang, and B. Schlikopf. Learning with hypergraphs: Clustering, classification, and embedding[C], *advances in neural information processing systems*. 2006: 1601-1608.

[14] G.B. Huang, M. Ramesh, T. Berg, and et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[R]. *Technical Report 07-49, University of Massachusetts*, Amherst, 2007.

[15] Z.G. Fan, J. Li, B. Wu, and et al. Local patterns constrained image histograms for image retrieval[C], *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008: 941-944.

[16] D. Gorisse, M. Cord, and F. Precioso. Locality-sensitive hashing for chi2 distance[J], *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(2): 402-409.

[17] M. Wang, X. Liu, and X. Wu. Visual Classification by-Hypergraph Modeling[J], *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(9): 2564-2574.

[18] B. Siddiquie, R.S. Feris, and L.S. Davis. Image ranking and retrieval based on multi-attribute queries[C], *Computer Vision and Pattern Recognition*, 2011: 801-808.

[19] Y. Li, R. Wang, H. Liu, and et al. Two birds, one stone: Jointly learning binary code for large-scale face image retrieval and attributes prediction[C], *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 3819-3827.