

# 融合距离度量学习和 SVM 的图像匹配算法

陈开志<sup>1,2</sup>, 乐承沛<sup>1,2</sup>, 钟尚平<sup>1,2</sup>

<sup>1</sup>(福州大学 数学与计算机科学学院 福州 350108)

<sup>2</sup>(福州大学 福建省网络计算与智能信息处理重点实验室 福州 350108)

E-mail: ckz@fzu.edu.cn

**摘要:** 目前度量学习方法通过有限样本数据学习得到新度量后,采用简单的分类器(如直接欧式距离计算)通常不能达到最佳分类效果. SVM 作为一种经典的分类器,具有优秀的线性和非线性分类能力,可以弥补距离度量学习方法的不足.对此,提出一种应用于图像匹配的融合距离度量学习和 SVM 的(DML-SVM)算法.首先,利用度量学习方法得到的线性变换矩阵,将样本变换到新的特征空间,降低特征各维度之间的相关性,调整特征各维度的权重;然后通过 SVM 对新特征进行线性或非线性的分类.通过在 LFW、Pubfig、ToyCars 三个图像数据库上的测试结果表明:融合方法的分类能力优于度量学习和 SVM 算法各自单独使用时的性能,且融合算法对训练样本数量具有很强的鲁棒性,即使只有少量训练样本(180个)时,融合算法仍然能具有较高的分类能力.

**关键词:** 机器学习;距离度量学习;SVM;图像匹配

中图分类号: TP181

文献标识码: A

文章编号: 1000-1220(2015)06-1353-05

## Fusion Method of Distance Metric Learning and SVM for Image Matching

CHEN Kai-zhi<sup>1,2</sup>, LE Cheng-pei<sup>1,2</sup>, ZHONG Shang-ping<sup>1,2</sup>

<sup>1</sup>(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

<sup>2</sup>(Fuzhou University Network Computing and Intelligent Information Processing Key Lab of Fujian Province, Fuzhou 350108, China)

**Abstract:** Generally, the distance metric learning method trained with limited samples can not achieve the best classification performance for the image matching by using a simple classification (such as Euclidean distance computing). SVM that is a classic classifier with excellent linear and nonlinear classification ability can make up for the lack of distance metric learning methods. This paper presents a fusion method of distance metric learning and SVM (DML-SVM) for image matching. First, the feature extracted from images is transformed into a new feature space by the transformation matrix obtained from DML, which reduces the correlation between each dimension of feature and optimizes the weight of each dimensions. Then the linear or non-linear SVM classifier is used to classify samples. The experimental results on the LFW, Pubfig and ToyCars databases show that the fusion method improve the recognition accuracy compared to single distance metric learning method or SVM, and the fusion method is robust to the number of training samples. Even if using only a small amount of training samples (180), the fusion method still has good classification performance.

**Key words:** machine learning; distance metric learning; SVM; image matching

## 1 引言

在模式识别的分类问题中,特征选择、相似度距离计算、分类器设计是三个基本的问题.早期的研究主要集中在特征选择和分类器的设计上,取的非常丰富的研究成果.而在相似度距离计算方面,通常采用一些固定的距离度量进行计算,如欧式距离、余弦距离、汉明距离等.但由于样本高维特征之间的量纲不统一、相关性,以及因分类的目标不同导致特征各维的权重不一致,使这种固定的距离度量在很多场景下不适用,只能通过不停的增强分类器的来弥补.近十年,学者们提出了一系列的距离度量学习方法(DML, Distance Metric Learning)通过将特征选择和相似度距离的选择这两个问题统一

起来考虑,利用样本进行监督或半监督学习,自适应的从特征中选择一个合适的度量进行分类.采用 DML 后,即使采用一些简单的分类器也可以显著地提高分类性能.这个框架<sup>[1,2]</sup>不仅将经典的 PCA、LDA 等方法,以及流形学习方法(ISO-MAP<sup>[3]</sup>、LLE<sup>[4]</sup>、LE<sup>[5]</sup>)纳入进来,还对它们有新的、更深入的认识,也产生了很多新的 DML 算法.因此,在机器学习、模式识别、数据挖掘等领域,距离度量学习方法受到越来越多的关注.

在过去的十年里,DML 成为了一大研究热点,取得了大量的研究成果.提出了很多好的度量学习方法[1].最早的度量学习文章是由 Xing 提出<sup>[6]</sup>,它的优化目标是寻找最佳的类似马氏距离计算中的变换矩阵  $W$ ,在满足同类型样本点之间距离之和小于一个固定值这个约束条件下,使得不同类

收稿日期: 2014-04-15 收修改稿日期: 2014-04-30 基金项目: 国家自然科学基金项目(61300025)资助;教育部博士点基金项目(20123514120013)资助;福州大学引进人才基金项目(022428)资助. 作者简介: 陈开志,男,1983年生,博士,讲师,研究方向为机器学习、多媒体安全技术;乐承沛,男,1989年生,硕士研究生,研究方向为机器学习、多媒体安全;钟尚平,男,1969年生,博士,教授,研究方向为机器学习、多媒体安全技术、隐写检测.

的样本点对的距离之和达到最大. 此后, 其它学者陆续提出了很多距离度量学习方法, 包括全局的<sup>[6]</sup>、局部的<sup>[7-9]</sup>学习方法. 单度量<sup>[7-9]</sup>、多度量<sup>[18, 19]</sup>的学习方法. 线性、非线性的学习方法. 其中比较经典的马氏距离度量方法有 LMNN<sup>[7]</sup>, IT-ML<sup>[8]</sup>, LDML<sup>[9]</sup>, CSML<sup>[12]</sup>, KISSME<sup>[11]</sup>, DML-egi<sup>[10]</sup>等方法. 这些方法中通过学习得到变换矩阵后, 都采用直接计算变换后的欧式距离进行分类. 这种分类虽然简单, 但前提是距离度量学习已经进行充分学习. 但实际上由于需要优化的变换矩阵含有大量的元素, 如 100 维的特征, 则变换矩阵有 10000 个元素 (100 × 100), 而训练样本点数量通常较少, 学习阶段很难满足充分训练的条件. 此外, 很多应用场景中, 要分类的样本特征变化是非线性的 (如人脸特征), 而距离度量学习方法通常只能捕捉线性变化. 虽然一些文献中提出的一些非线性的 DML 方法<sup>[13-18]</sup>, 如  $\chi^2$ -LMNN、GB-LMNN<sup>[13]</sup>, 与核方法的结合<sup>[15, 16]</sup>等, 但这些方法的共同问题是由于要学习的变换矩阵元素过多, 导致优化过程异常复杂而且敏感. 只能因此本文希望通过采用合适的分类器来弥补度量学习方法的不足, 提高分类性能.

模式识别中, 分类器的研究一直是个热点. 其中, SVM (Support Vector Machine) [20-22] 因其优异的性能, 成为目前应用最广、研究最多的分类器之一. SVM 根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷, 以期获得最好的推广能力 (或称泛化能力). 它在解决小样本、非线性及高维模式识别中表现出许多特有的优势. SVM 分类器对训练样本的数量没有太大的要求, 即使只有少量的训练样本也能获得很好的分类性能. 而且 SVM 不仅可以对线性情况进行分类, 通过和核方法结合还能很好的处理非线性的情况. 这些优点正好可以弥补 DML 方法的不足. 而对于样本高维特征之间的量纲不统一、相关性, 以及因分类的目标不同导致特征各维的权重不一致时, SVM 方法表现较差, 而这个正是 DML 所擅长的. 从 DML 方法与 SVM 方法的优缺点可知, 两者可以互相补短. 因此, 本文研究将距离度量学习方法和 SVM 融合使用的算法.

本文余下的组织结构如下: 第二节介绍距离度量学习方法和 SVM 的线性和非线性的分类方法, 分析优缺点; 第三节介绍本文的方法; 第四节给出实验结果并进行分析; 最后是对本文工作的总结.

## 2 相关知识

### 2.1 距离度量学习方法

模式分类中进行距离计算时, 采用固定的距离度量常面临样本特征量纲不统一的问题. 如用人的两个特征 (体重 (kg) 和身高 (m)) 去分类这个人适不适合打篮球. 实际情况应该是身高要占主导地位, 但如果直接用欧式距离计算, 体重在最终距离度量中占主要地位, 将导致错误的分类结果. 而且对不同的应用, 不同的特征属性可能不一样. 例如要去分类适不适合拔河, 此时体重应该占主导地位, 即分类的目标不同导致特征各维的权重应该不同. 此外, 特征之间通常有着一定的相关性 (如通常身高越高、体重越重, 但又不是完全成正比), 高维时甚至有很复杂的相关性, 包括线性和非线性的. 这些问

题融合在一起, 导致在很多应用中不适合采用固定度量来判断样本相似性. 因此, 根据特定问题从样本数据中学习一个合适的度量来衡量相似度距离更符合问题的本质. DML 方法即符合这种特性, 它借鉴了马氏距离的思想. 样本  $x$  和  $y$  的马氏距离定义如下:

$$d_M(x, y) = \sqrt{(x - y)^T M (x - y)} \quad (1)$$

其中  $M \in R^{d \times d}$  为变量  $X$  与  $Y$  的协方差阵的逆矩阵, 为对称半正定矩阵. 如果  $M$  是一个单位矩阵  $I$ , 则马氏距离转变为欧式距离. 如果  $M$  不是一个单位矩阵, 假设它的秩为  $r$ , 则马氏矩阵  $M$  可以采用 Cholesky 分解为  $M = W^T \times W$ , 其中  $W \in R^{r \times d}$ , 则可以重写马氏距离的表达式如下:

$$\begin{aligned} d_M(x, y) &= \sqrt{(x - y)^T W^T W (x - y)} \\ &= \sqrt{(Wx - Wy)^T (Wx - Wy)} = \|Wx - Wy\|_2 \end{aligned} \quad (2)$$

其中  $\|\cdot\|_2$  是欧氏距离的标准范式. 从上式可知马氏矩阵实际上对样本空间的进行一个线性变换, 然后在变换后的空间计算欧式距离.

虽然马氏距离是经过归一化、去相关处理后的距离度量, 能解决量纲不一致、线性相关等问题, 但不能解决非线性变化和各维的权重配置的问题. DML 方法借鉴了马氏距离的思想, 但其中矩阵  $M$  不再是协方差阵, 而是通过构造优化目标, 利用样本数据进行监督/半监督学习获得最优的变换矩阵  $M$ , 即构造出一个适合特定应用的距离度量函数. 如最早由 Xing 提出的全局算法, 优化的目标是使变换后的同类数据点收缩, 使不同类数据点分离. 用优化函数表示即约束某一类距离的条件下, 最小化或最大化另一类的距离和.

$$\begin{aligned} \min_M \quad & \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_M^2 \\ \text{s.t.} \quad & \sum_{(x_i, x_j) \in D} \|x_i - x_j\|_M^2 \geq 1, M \geq 0 \\ \text{或} \quad & \max_M \sum_{(x_i, x_j) \in D} \|x_i - x_j\|_M^2 \\ \text{s.t.} \quad & \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_M^2 \leq 1, M \geq 0 \end{aligned} \quad (3)$$

这种通过学习获得的度量除了具有马氏距离的功能外, 还能根据标记的训练样本进行学习, 合理配置各维的权重. 通过构造特定的优化目标, 还可能处理特征非线性变化的问题. 这些问题的处理, 都有助于使学习到的距离度量最大程度的满足最终的分类目的.

DML 方法通常涉及到凸优化问题的求解. 在高维中由于矩阵  $M \in R^{d \times d}$  的元素个数是特征维度的平方, 直接求解消耗很大, 且需要大量的样本数据进行训练. 很多应用场景中无法收集足够的数据进行训练, 导致训练不充分. 此外, 如果要处理非线性变化, 只能定义一些特定的非线性度量, 如<sup>[13-18]</sup>, 再结合凸优化进行处理. 这种方法要求已知数据非线性分布的先验知识, 才能定义一个合理的非线性度量. 在实际情况中很难满足. 如果采用核方法, 同时优化矩阵和核方法参数将使系统变得异常复杂, 且不稳定.

### 2.2 SVM 方法

支持向量机是从线性可分情况下的最佳超平面发展而来. 对于一组带有类别标记的训练样本集  $(x, y)$ , 若超平面

$w \cdot x + b = 0$  能将样本正确分为两类,并使两类样本到超平面最小距离之和最大,则该超平面就是最佳超平面。SVM 通过求解下面的优化问题得到。

$$\min \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \quad (4)$$

$$s. t. \quad y_i (\tilde{\omega} \cdot x_i) + b \geq 1 - \xi_i, i = 1, \dots, l$$

$$\xi_i \geq 0, i = 1, \dots, l$$

其中  $\xi_i$  为误差项,  $C$  为惩罚系数。利用 Lagrange 乘子法可以把上述求解最佳超平面的问题转化为对偶问题。由此得到决策函数

$$f(x) = \text{sign} \left( \sum_{i=1}^l a_i y_i (x_i \cdot x) + b \right) \quad (5)$$

SVM 方法还能很方便的和核方法结合,用于处理非线性的模式识别问题。对于线性不可分情形,先利用一个非线性映射将数据映射到一个高维的特征空间实现线性分类。相应的决策函数变为

$$f(x) = \text{sign} \left( \sum_{i=1}^l a_i y_i k(x_i, x) + b \right) \quad (6)$$

其中  $k(\cdot)$  称为核函数,核函数的选取应使其为特征空间的一个内积。常用的核函数有:多项式核函数、Gauss 径向基核函数、Sigmoid 核函数。

但 SVM 的问题是它计算距离时,它们都隐含假定待分析样本向量的各维特征对分类的贡献均匀,不考虑各个特征对分类的不同影响。核函数是 SVM 研究的核心问题,常用核函数是通过描述样本相似性的内积或距离来定义的,而内积或距离是根据样本的所有特征计算的。在这些特征中,有些特征与分类强相关,有些特征与分类弱相关,还有些特征与分类不相关。这样,核函数的计算可能会被一些弱相关或不相关的特征所支配,从而影响分类器的分类性能。虽然一些文献针对这些问题提出了相应的改进算法,如特征加权支持向量机 FWSVM。

### 3 基于距离度量学习的 SVM 算法

#### 3.1 为什么融合?

从上一节对距离度量学习方法和 SVM 方法的优缺点分析可知,对于样本高维特征之间的量纲不统一、相关性,以及因分类的目标不同导致特征各维的权重不一致,DML 方法可以很好的处理,但它对于数据样本训练数量要求很高,且对非线性特征处理不好。SVM 分类器对训练样本的数量没有很好的要求,即使只有少量的训练样本也能获得很好的分类性能。而且 SVM 不仅可以对线性情况进行分类,通过和核方法的结合还能很好的处理非线性的情况。这些优点可以弥补 DML 方法的不足。SVM 方法对处理样本高维特征之间的量纲不统一、高度相关性,以及因分类的目标不同导致特征各维的权重不一致这些方面表现较差,而 DML 方法可以很好的处理这些问题。鉴于 DML 方法和 SVM 方法各自的优缺点,二者融合是可以取长补短。因此,本文研究将 DML 方法和 SVM 融合使用的算法。

#### 3.2 融合方法 DML-SVM

根据度量学习的目标  $L(M)$  进行最优化处理,得到一个

度量矩阵  $M$  或者是一个线性变换矩阵  $W$ 。如果学习得到的是度量矩阵  $M$ ,可以利用 Cholesky 分解得到线性变换矩阵  $W$ 。Cholesky 分解如下:

$$M = W^T W \quad (M \in R^{d \times d}, W \in R^{r \times d}) \quad (7)$$

$r$  对应的是度量矩阵  $M$  的秩,并且  $r \leq d$ ,所以如果要求在过度量矩阵线性变换之后,样本可以达到降维的效果,就需要对矩阵  $M$  进行一个低秩的规则化,使  $M$  保持低秩特性。经过度量矩阵的线性变换之后,计算两点的距离向量,

$$\Phi_{ij} = Wx_i - Wx_j \quad (8)$$

然后,在新的距离空间中,对 SVM 方法进行训练,获得分类器

$$f(\Phi) = \text{sign}(a_i y_i (\Phi_i \cdot \Phi) + b) \quad (9)$$

最终用该分类器对样本进行分类。

DML 与 SVM 融合具体方法如下:

训练方法:

输入: 成对样本训练集  $S \cup D$

输出: 马氏矩阵  $M = W^T W$  (或者线性变换矩阵  $W$ ) 和 SVM 分类器  $f(\Phi)$

1. 根据训练集  $S \cup D$ , 利用具体的一个马氏度量学习方法学习得到马氏矩阵  $M$ , 对其进行 Cholesky 分解, 得线性变换矩阵  $W$ ;
2. 利用步骤 1 的线性变换矩阵  $W$  将训练样本集线性映射到另一个空间;
3. 计算距离向量  $\Phi_{ij}$ , 形成新样本集  $(\Phi_{ij}, y_{ij})$ , 用于 SVM 训练
4. 用训练 SVM 分类器, 得到参数集  $(a_i, b)$ , 即分类器  $f(\Phi)$

分类方法:

输入: 测试样本  $T$ , 马氏矩阵  $M = W^T W$  (或者线性变换矩阵  $W$ ) 和 SVM 分类器  $f(\Phi)$

输出: 分类结果  $Y$

1. 利用步骤 1 的线性变换矩阵  $W$  将测试样本映射到另一个空间;
2. 计算样本距离向量  $\Phi$ ,
3. 用 SVM 分类器  $f(\Phi)$ , 计算分类结果  $Y$

以上方法中, SVM 可以采用线性形式的分类器,也可采用非线性形式的分类器。

### 4 实验结果

本文实验的目标有两个: 第一个是为了证明距离度量学习方法与 SVM 方法的融合是能够提高分类的准确率; 第二个是为了证明融合算法对训练样本的数量具有更好的鲁棒性。

为验证这两点,本文采用了 4 个经典的距离度量学习方法(KISSME, ITML, LMNN, LDML)与 SVM 方法相融合,通过比较融合后算法和各自独立使用时算法的各项性能指标来实现以上两个目标。其中度量学习的代码采用了 KISSME 作者提供的源码<sup>[23]</sup>,它集成了上面的四个距离度量学习方法。SVM 分类器的 libSVM 工具包<sup>[22]</sup>进行处理。

实验的环境是 windows7 下的 matlab2013a,采用的计算机主要配置为: intel i5-2310 四核 CPU, 频率 2.9G, 8G 内存。本文所用的测试数据库和文献[11]相同,都采用 LFW<sup>[24]</sup>、Pubfig<sup>[25]</sup>、Toycars<sup>[26]</sup>三个数据库。其中 LFW 和 Pubfig 属于人脸数据库,都是非约束环境下采集的人脸图片; Toycars 是汽车图片数据库。

#### 4.1 算法分类性能

#### 4.1.1 LFW 数据库实验结果

LFW 数据库包含了 5749 个人的 13233 张图片并且每张图片都已经确认了身份信息. 在这个数据库上有两个数据集: 一是把 5749 个人的图片直接分为训练集和测试集; 二是把这些图片分为 10 个子集, 每个子集含有相似的 300 对图片和不相似的 300 对图片, 并且每个人的图片不能出现在多个子集中. 本文采用的是第二个数据集, 并且遵从“严格约束”, 即不利用样本的身份信息, 只能利用给定的样本对是否是同一人的约束信息.

实验中, 人脸的特征表示采用 Guillaumin 提供的 sift 特征. 该特征是在人脸上检测出的九个点在三个尺度下提取的 128 维特征, 所以总维数是  $9 \times 3 \times 128 = 3456$  维. 由于维数过大, 先采用 PCA 方法降到 100 维, 再采用对应的分类方法处理. 测试人脸识别的正确率, 结果如表 1 所示.

表 1 LFW 数据库上的识别率

Table 1 Recognition accuracy in LFW database

	SVM 单独使用 (线性/非线性)	DML 单独使用	DML-SVM (线性/非线性)
ITML		78.90%	79.70% / 80.35%
LMNN	72.71% / 73.57	68.20%	70.27% / 70.50%
LDML		79.60%	81.16% / 82.00%
KISSME		80.50%	84.03% / 84.93%

从表 1 的结果可以看出, 除 LDML 外, 其他 DML 与 SVM 方法两者的融合的正确率都要比各自单独使用高 1% ~ 4%. 其中, KISSME 方法与 SVM 方法的融合达到了最好的效果, 线性 DML-SVM 比 KISSME 方法的正确率提高了 3.53%, 比 SVM 方法的正确率提高了 11.32%. 采用非线性 DML-SVM 还可以再提高 0.9%. 虽然 LDML 方法和线性 SVM 融合时性能没有提高, 但其采用非线性 SVM 时, 性能提高了 2.4%.

#### 4.1.2 Pubfig 数据库实验结果

Pubfig 人脸数据库采集了因特网上的 200 个人的 58797 张图片. 这个数据集被分成了两个部分, 一个是包含了 60 个人的 Development Set, 一个是包含了剩下的 140 个人的 Evaluation set. 第二数据集是用来测试算法的性能的, 所以用的是第二数据集. 这个数据集被分成了 10 个子集交叉验证, 每个

表 2 Pubfig 数据库上的识别率

Table 2 Recognition accuracy in pubfig database

	SVM 单独使用 (线性/非线性)	DML 单独使用	DML-SVM (线性/非线性)
ITML		69.10%	76.16% / 76.91%
LMNN	74.03% /	73.20%	74.56% / 75.07%
LDML	74.45%	77.60%	78.45% / 78.72%
KISSME		77.60%	79.54% / 79.72%

子集包含有 1000 对相似的图片, 1000 对不相似的图片. 每个人、每张图片都是不相交地分布在每个子集中. Pubfig 数据库里的图片特征是有一系列的命名的属性构成的, 比如: 年龄、种族、性别、头发等等, 一共 73 维. 该属性值是大于零的表示该属性特征是有出现的, 属性值小于零的表示该属性特征

是没有出现的.

Pubfig 数据库的实验结果如表 2 所示, 从结果可知, 与 KISSME、ITML、LMNN 三个各自的马氏度量方法的正确率相比, SVM 与它们的融合可以提升 1% ~ 7.5% 的正确率, 特别是与 ITML 的度量方法的融合, 比单独度量学习方法的正确率提高约 7%, 比单独使用 SVM 方法的正确率也提高了 2%.

#### 4.1.3 Toycars 数据库实验结果

ToyCars 数据集是测试之前没有见过的汽车模型的图片的相似度. 该数据集包含了 14 种不同的汽车和卡车的 256 张图片, 它展示出了不同的姿态、光照和背景. 这个数据集采用的是 2 个子集交叉验证, 在一个子集中包含了 7 种汽车图片, 总共有 1185 对的相似图片对, 7330 对的不相似图片对. 同样剩下的 7 种汽车图片对在另一个子集中. 对于图片的特征表示是采用了文章 [1] 的做法, 将图片分割成没有相互覆盖的  $30 \times 30$  的图片块, 提取 LBP 特征, 最后把各个局部的特征串联成全局的特征, 并用 PCA 方法降到 50 维.

表 3 Toy cars 数据库上的识别率

Table 3 Recognition accuracy in toy cars database

	SVM 单独使用 (线性/非线性)	DML 单独使用	DML-SVM (线性/非线性)
ITML		70.70%	85.85% / 85.97%
LMNN	84.37% /	83.50%	91.27% / 88.32%
LDML	85.86%	71.60%	85.80% / 85.97%
KISSME		93.40%	94.03% / 94.47%

表 3 给出了在 Toy cars 数据库上的实验结果, 从中可知, 融合后的方法分类正确率都得到提升. 其中 LMNN 和线性 SVM 融合后, 正确率提升了 7.65%.

综合三个数据库的实验结果, 除个别度量学习方法在某个数据库下融合算法的性能没有提高外, 大多数性能都提高了 1% ~ 7%, 表明融合算法有效性. 四个融合算法中, KISSME + SVM 在三个数据库都取的最佳的性能. 此外, 采用非线性 SVM 方法时, 无论是单独使用还是在融合算法中使用都只比线性 SVM 有微小的提高 (1% 以内), 这主要是由于训练样本较多和特征高维时, 非线性方法无法体现出优势.

#### 4.2 算法时间效率比较

本文采用了 4 个经典的距离度量学习方法 (KISSME、ITML、LMNN、LDML) 与 SVM 方法相融合. 后三种方法都是对不同的学习目标函数, 训练时采用凸优化技术, 通过迭代计算来寻找最优的度量矩阵  $M$ , 非常耗时. 而 KISSME 在特征正态分布的假设下, 巧妙地将度量学习方法中的优化迭代训练转化成简单的统计计算来近似一个最优的度量矩阵, 大大降低了样本训练时间, 解决了样本数据量过大时距离度量学习耗时严重的问题. 如文献 [11] 中提供的数据: 用 LFW 数据库中的 4500 对图片训练, 后三种方法训练时间分别约为 24、1198、307 秒, 而 KISSME 的训练方法只有 0.05 秒. 此外, 四种 DML 算法中, KISSME 分类性能最好, 所以这里以最实用的 KISSME 作为 DML 算法代表, 比较融合前后算法的时间性能, 如下页表 4 所示. 从表中可知, SVM 算法、Kissme 以及融合算法的训练都很快, 在 LFW 和 toy cars 数据库中都不到 0.1 秒, PUBFIG 中最高也只要 0.55 秒, 说明融合算法 Kissme + 线性 SVM 具有很高的时间效率.

### 4.3 算法对训练样本数量的鲁棒性

从以上算法分类的实验结果可知,四种 DML 算法中, KISSME 分类性能最好,训练速度最快.这里以 KISSME 算法作为 DML 算法代表,在 LFW 数据库上测试独立算法和融合算法对训练样本数量的鲁棒性.结果如表所示.从表中可知,

表 4 融合前后算法的训练时间对比

Table 4 Training time of fusion method

方法	Lfw	PUBFIG	Toycars
线性 SVM	0.083s	0.550s	0.067s
Kissme	0.050s	0.070s	0.040s
Kissme + 线性 SVM	0.072s	0.294s	0.060s

随着训练样本数从 2700 降低到 180, KISSME 分类正确率下降了 18%, 而融合后的线性 SVM + KISSME 下降 7%, 而非线性 SVM + KISSME 只下降 2%. 但如果训练样本继续降低到

表 5 不同训练样本数量下的分类正确率

Table 5 Recognition accuracy with different training size

训练样本数量	KISSME	线性 SVM + KISSME	非线性 SVM + KISSME
2700	79.10%	83.75%	84.20%
1800	77.60%	83.55%	84.30%
900	75.60%	82.55%	84.32%
450	73.47%	79.93%	83.28%
270	68.44%	78.87%	83.38%
180	62.66%	76.78%	82.38%
90	61.86%	75.02%	80.07%
45	59.68%	72.35%	74.18%

45 时, 融合算法性能也开始恶化, 但分类能力仍然优于单纯的 KISSME 算法. 以上结果说明融合算法中 SVM 可提高 KISSME 算法对训练样本数量的鲁棒性; 在训练样本数量较少时, 非线性 SVM 方法具有更好的性能.

## 5 总 结

本文针对度量学习方法和 SVM 方法的优势和不足, 提出了基于度量学习方法的 SVM 方法. 为了验证融合方法的分类性能和对训练样本数量的鲁棒性, 本文采用了四个经典的度量学习方法, 在三个具有挑战性和大规模的数据库上进行了实验验证, 包括 LFW、Pubfig、Toycars. 实验结果表明该融合方法是可以提升分类性能, 并且降低度量学习算法对训练样本数量的要求.

### References:

- [1] Kulis B. Metric learning: a survey [J]. Machine Learning, 2012, 5(4): 287-364.
- [2] Yang L. An overview of distance metric learning [C]. IEEE Conference on Computer Vision and Pattern Recognition Conference (CVPR), 2007.
- [3] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(5500): 2319-2323.
- [4] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.
- [5] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. Neural Computation, 2003, 15(6): 1373-1396.
- [6] Xing E P, Ng A Y, Jordan M I, et al. Distance metric learning with application to clustering with side-information [C]. Advances in Neural Information Processing Systems (NIPS), 2003: 521-528.
- [7] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification [C]. Advances in Neural Information Processing Systems (NIPS), 2006, 18: 1473.
- [8] Davis J V, Kulis B, Jain P, et al. Information-theoretic metric learning [C]. Proceedings of the 24th International Conference on Machine Learning (ICML), ACM, 2007: 209-216.
- [9] Guillaumin M, Verbeek J, Schmid C. Is that you? Metric learning approaches for face identification [C]. IEEE 12th International Conference on Computer Vision (ICCV), 2009: 498-505.
- [10] Ying Y, Li P. Distance metric learning with eigenvalue optimization [J]. The Journal of Machine Learning Research, 2012, 13: 1-26.
- [11] Kostinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012: 2288-2295.
- [12] Nguyen H, Bai L. Cosine similarity metric learning for face verification [C]. Asian Conference on Computer Vision (ACCV), 2011: 709-720.
- [13] Kédem D, Tyree S, Weinberger K Q, et al. Lanckriet, nonlinear metric learning [C]. Advances in Neural Information Processing Systems (NIPS), 2012: 2582-2590.
- [14] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively with application to face verification [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, 1: 539-546.
- [15] Jain P, Kulis B, Davis J V, et al. Metric and kernel learning using a linear transformation [J]. Journal of Machine Learning Research, 2012, 13: 519-547.
- [16] He Y, Chen W, Chen Y. Kernel density metric learning [R]. Technical Report, Washington University in St. Louis, 2013.
- [17] Torresani L, Lee K. Large margin component analysis [C]. Advances in Neural Information Processing Systems (NIPS), 2007: 1385-1392.
- [18] Weinberger K Q, Saul L K. Fast solvers and efficient implementations for distance metric learning [C]. Proceedings of the 25th International Conference on Machine Learning (ICML), ACM, 2008: 1160-1167.
- [19] Cui Zhen, Li Wen, Xu Dong, et al. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild [C]. IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR), 2013.
- [20] Vapnik V. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995: 91-188.
- [21] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods [M]. Cambridge University Press, 2000: 47-98.
- [22] Chang C C, Lin C J. LIBSVM: a library for support vector machines [C]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.
- [23] <http://lrs.icg.tugraz.at/research/kissme/>, 2014.
- [24] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments [R]. University of Massachusetts, Amherst, Technical Report, 2007: 7-49.
- [25] Kumar N, Berg A C, Belhumeur P N, et al. Attribute and simile classifiers for face verification [C]. IEEE 12th International Conference on Computer Vision (ICCV), 2009: 365-372.
- [26] Nowak E, Jurie F. Learning visual similarity measures for comparing never seen objects [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007: 1-8.