

基于结构化稀疏投影的多视图特征提取框架

江帆¹, 田青²

(1. 南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106;

2. 南京信息工程大学 计算机与软件学院, 江苏 南京 210044)

摘要: 多视图学习通常基于两个重要原则: 一致性原则和互补性原则。视图的一致性源自所有视图间的共享信息; 而视图的互补性源自不同视图的特有信息。然而, 现有多视图学习算法通常只关注其中的一种, 以致得到的模型性能非最优。为此, 提出一种新的多视图特征提取算法框架, 称为多重结构化稀疏投影(MSSP), 其能同时提取视图的公共和特有信息。MSSP的目标函数包含两个部分: 融合投影的判别项和联合投影矩阵的结构化稀疏正则化项。通过以上建模方式, 不同视图的一致和互补信息同时得到了融合利用。此外, 给出所建模型基于在Stiefel流形上的梯度下降的求解算法, 获得局部最优解, 并分析推测了模型参数随投影维数的变化趋势。相关对比实验验证了MSSP的建模有效性和参数的变化趋势。

关键词: 一致性; 互补性; 多视图学习; 特征提取; 结构稀疏; Stiefel流形

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2019)03-0006-06

doi: 10.3969/j.issn.1673-629X.2019.03.002

A Framework of Multi-view Feature Extraction Based on Multiple Structured Sparsity Projection

JIANG Fan¹, TIAN Qing²

(1. School of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China;

2. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Multi-view learning is usually based on two important principles: consistency and complementarity. Consistency derives from the shared information between views and complementarity derives from the unique information of different views. However, the existing multi-view learning algorithms usually focus on only one of them, so that the resulting model performance is not optimal. For this, we propose a new multi-view feature extraction algorithm framework called multiple structured sparsity projection (MSSP), which can simultaneously extract the common and unique information of the view. The objective function of MSSP contains two parts: fusion projection discriminant and structured sparsity regularization. Then the consistent and complementary information of different views can be fused simultaneously. In addition, we also give an algorithm based on the gradient decent over the Stiefel manifold to obtain a local optimal solution and analyze how the hyper-parameter change with the dimension of the projection. The experiment verifies the effectiveness of MSSP and the trend of the hyper-parameter.

Key words: consistency; complementarity; multi-view learning; feature extraction; structured sparsity; Stiefel manifold

1 概述

在实际的机器学习任务中, 目标常常以多视图数据描述。例如, Web 页面可以用文本、图像和链接等组件来共同描述, 一个组件的数据是页面的一个视图。每一种视图都有其特殊的结构和独特的概念; 而不同的视图又是相互关联的, 因为它们描述的对象是相同

的。多视图数据的这两种性质蕴含着多视图学习的两个原则: 互补性和一致性^[1]。一致性源自不同视图之间的相关性, 互补性源自每个视图的独有信息。根据一致性和互补性, 多视图降维算法可分为三类。第一类是只关注一致性, 如凸子空间表示学习(convex subspace representation learning, CSRL)^[2]和典型相关分析

收稿日期: 2018-03-21

修回日期: 2018-07-25

网络出版时间: 2018-12-19

基金项目: 国家自然科学基金(61702273); 江苏省自然科学基金(BK20170956)

作者简介: 江帆(1985-), 男, 硕士研究生, CCF 会员(54494G), 研究方向为模式识别与机器学习。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181219.1510.010.html>

(canonical correlation analysis, CCA)^[3-4]。第二类是只侧重于互补性,如多视图判别分析(multi-view discriminant analysis, MVDA)^[5]和集成流形正则化稀疏低秩逼近(ensemble manifold regularized sparse low-rank approximation, EMRLRA)^[6]。最后一类是同时关注两种原则,如基于结构化稀疏的分解隐空间学习(factorized latent spaces with structured sparsity, FLSS)^[7]和部分共享的隐因子学习(partially shared latent factor learning, PSFL)^[8]。

而根据多视图数据的基本假设,多视图降维算法可以分为两类。第一种假设是低维表示只包含原始数据信息的一部分。不同的算法根据不同偏好提取相应信息,如判别信息和流形信息。投影学习模型(projection learning model)^[9],如多视图典型相关分析(multi-view canonical correlation analysis, MVCCA)^[10]和 MV-DA,都是基于这种假设。另一种假设是不同的高维视图的数据是由相同的低维数据生成的。基于隐变量生成模型(latent variable generation model)^[11-12]的算法,如高斯过程(Gaussian process, GP)^[13-14]和 CSRL,都是基于这种假设。

综合考虑两个原则和两个假设,多视图降维算法可以更细致地分为六类。表1为六类算法及其示例。

表1 六类算法及其示例

Model\principle	Consistency	Complementarity	Both
Generation	CSRL, GP	EMRLRA	FLS, PSFL
Projection	CCA, MVCCA	MVDA	φ

现今尚没有同时关注两个原则的基于投影模型的多视图降维算法。为了解决这个问题,提出了一种新的基于结构化稀疏投影的监督型多视图特征提取框架(MSSP)。框架由两部分组成。一部分是融合特征的判别损失函数。大多数单视图投影学习算法的损失函数都可以实例化这部分,文中选择线性判别分析(linear discriminant analysis, LDA)^[15]和度量学习^[16]。另一部分是联合投影矩阵的多重结构化稀疏正则化。

文中主要贡献如下:

(1) 在向多视图数据上拓展单视图算法时,一个简单的方案是将所有视图的特征拼接成一个单一的视图数据。然而,这种方案不但忽略了不同视图潜在的相关性,也缺乏物理意义。MSSP的正则化项解决了这个问题。

(2) MSSP是同时关注一致性和互补性的基于投影模型的多视图降维算法,能够同时提取融合视图的共享信息和独有信息,避免信息缺失和信息冗余的问题。

(3) 定义了一个 M 范数作为正则化项,以惩罚联

合投影矩阵的多重结构化稀疏程度,其中联合投影矩阵约束为列正交矩阵。然后利用在 Stiefel 流形上的梯度下降法,对框架优化求解,得到一个局部最优解。

(4) 分析了 MSSP 的目标函数的数学性质并推测了超参数的选择与投影维数的关系,以缩小超参数的选择范围,降低模型训练的代价。

2 MSSP 原理及相关工作

2.1 MSSP 投影融合原理

图1展示了并行信息融合模型^[17]中一致性和互补性的简单解释。其中每种颜色代表完整信息的一部分,视图1只拥有红色与绿色信息,视图2只拥有黄色与绿色信息,两个视图融合后得到三种颜色的完整信息。由于视图2缺少红色信息,故在提取红色信息时,视图2的投影必须为零,即视图2的投影向量为零向量,否则融合的红色信息将不纯。提取黄色信息与红色信息是相似的。红色信息和黄色信息分别为视图1和视图2独有的,体现互补性。与红色和黄色信息不同,绿色信息是两种视图的共享部分,体现一致性。如果两个视图联合提取绿色信息,则更有效、更准确。这意味着两个视图的绿色信息投影向量应该都是非零的。综上,可以发现每个视图的投影矩阵都必须是列组稀疏,如图中 W_1 和 W_2 所示,而联合投影矩阵 $[W_1^T, W_2^T]$ 的稀疏形式则称为多重结构化稀疏。

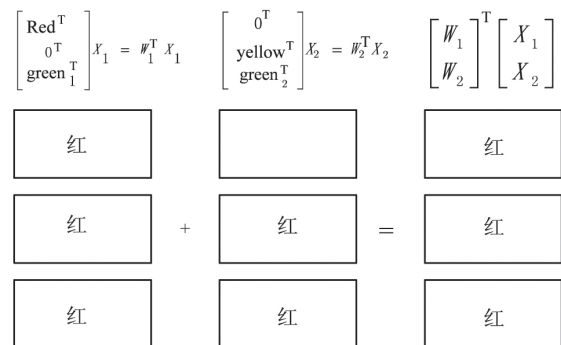


图1 并行信息融合模型的一致性和互补性解释

2.2 稀疏学习

在训练数据是稀疏的或者模型有特定的偏好时,模型的解决方案(向量或矩阵)需要拥有特定的稀疏结构。因 L_1 范数具有包容性、凸性和强大的理论保证等优势,大多数现有的稀疏学习都是基于其正则化的变种^[18]。1996年, Tibshirani 提出了一种稀疏特征选择算法 Lasso (The least absolute shrinkage and selection operator), 可以自适应地选择有用的特征。其形式如下:

$$\min_{\beta} \|L - X\beta\|_2^2 + \lambda \|\beta\|_1$$

其中, $X \in R^{p \times q}$ 为数据矩阵; $L \in R^p$ 为类标签向量; $\lambda > 0$ 为控制稀疏程度的超参数; $\beta \in R^q$ 为选择特

征的稀疏向量。

Group Lasso 是 Lasso 的一个变种。Lasso 将单个特征作为单位而 Group Lasso 将变量组作为单位进行选择。Group Lasso 的形式如下:

$$\min_{\beta} \|L - X\beta\|_2^2 + \lambda \sum_{i=1}^l \|\beta_{\theta_i}\|_2$$

其中, $\theta_i \in N_q$, $\prod_{i=1}^l \theta_i = N_q$, β_{θ_i} 是 β 中以 θ_i 中元素为下标的元素组成的一个子向量。

Lasso 和 Group Lasso 能够覆盖向量的大多数稀疏形式。对于矩阵, 惩罚项通常为 $L_{2,1}$ 范数或 $L_{\infty,1}$ 范数。 $L_{2,1}$ ^[19] 和 $L_{\infty,1}$ 的形式如下:

$$\|Z\|_{2,1} = \sum_{i=1}^d \|Z^i\|_2, \|Z\|_{\infty,1} = \sum_{i=1}^d \|Z^i\|_{\infty}$$

其中, $Z \in R^{s \times d}$, Z^i 表示矩阵 Z 的第 i 列。

2.3 Stiefel 流形

正交是一种常见的矩阵结构, 通常用于减少模型的复杂度或应用于可视化。在文献[20]中, 线性降维算法的解被约束为正交矩阵, 算法被看作矩阵优化问题。应用在 Stiefel 流形上的梯度下降法可以得到一个局部最优解。Stiefel 流形 $O^{d \times r}$ 是列正交矩阵的集合, 可以表示为 $O^{d \times r} = \{M \in R^{d \times r} \mid M^T M = I\}$, 其中 I 是 $r \times r$ 的单位矩阵。 $O^{d \times r}$ 是 $R^{d \times r}$ 的子流形嵌入, 是有界的闭集。切空间是流形在某一特定点上的线性逼近, 对于理解任何流形几何, 特别是矩阵流形的优化至关重要。首先定义流形 $O^{d \times r}$ 上的一条曲线, 一个光滑的映射 $\gamma(\cdot) : R^{d \times r} \rightarrow O^{d \times r}$, 进而定义切空间为:

$$T_M = \{\dot{\gamma}(\cdot) : \gamma(\cdot) \text{ is a curve on } O^{d \times r} \text{ with } \gamma(0) = M\}$$

其中 $\dot{\gamma}$ 为 $d\gamma(t)/dt$ 。粗略地讲, T_M 是在点 M 上沿着流形变化的所有方向的集合, 切空间有助于理解正交矩阵在梯度下降法中如何更新。

3 MSSP 框架的建立

3.1 符号和相关定义

给定包含 c 类的多视图数据, 第 i 个视图的数据矩阵为 $X_i \in R^{d_i \times n}$, 其投影矩阵为 $W_i \in R^{d_i \times d}$, 其中 d_i 为第 i ($i = 1, 2, \dots, p$) 个数据矩阵的特征维数, d 是给定的投影维数。联合数据矩阵为 $X = [X_1^T, X_2^T, \dots, X_p^T]^T = [X^1, X^2, \dots, X^p]$, 其中 X^j 表示第 j 类数据。联合投影矩阵为 $W = [W_1^T, W_2^T, \dots, W_p^T]^T$ 。定义 L_M 范数为

$$\|W\|_M = \sum_{i=1}^p \|W_i\|_{2,1}。各视图投影以并行方式^[17]$$

$$\text{融合 } Y = \sum_{i=1}^p W_i^T X_i = W^T X。$$

度量学习^[16]中, 融合特征的类内和类间距离矩阵如下:

$$S_W = \sum (n_k X^k X^{kT} - X^k \mathbf{1}_{n_k} \mathbf{1}_{n_k}^T X^{kT}) \quad (1)$$

$$S_B = n X X^T - X \mathbf{1}_n \mathbf{1}_n^T X^T - S_W \quad (2)$$

LDA 中, 融合特征的类内和类间散度矩阵如下:

$$\tilde{S}_W = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_j^i - u^i) (x_j^i - u^i)^T \quad (3)$$

$$\tilde{S}_B = \sum_{i=1}^c n_i (u^i - u) (u^i - u)^T \quad (4)$$

其中, x_j^i 为 X^i 的第 j 个样本; u^i 为 X^i 的均值; n_i 为 X^i 的样本数; u 为 X 的均值。

3.2 MSSP 模型

MSSP 的基本框架的抽象形式如下, 大多数的机器学习算法都可以抽象为这种形式:

$$\begin{aligned} \min_W f(W) &= \text{loss}(W; X) + \lambda \mathfrak{R}(W) \\ \text{s. t. } W &\in \Psi \end{aligned} \quad (5)$$

其中, $\text{loss}(W; X)$ 为数据 X 上关于 W 的损失函数; $\mathfrak{R}(W)$ 为 W 的正则化; λ 为平衡损失与正则化的超参数; Ψ 为 W 的约束空间。

MSSP 以 $\|W\|_M$ 和 Stiefel 流形实例化 $\mathfrak{R}(W)$ 和 Ψ 。文中用 LDA 的瑞利商形式和度量学习来实例化损失函数, 因此有 MSSP_LDA 如下:

$$\begin{aligned} \min_W f(W) &= \frac{\text{tr}(W^T \tilde{S}_W W)}{\text{tr}(W^T \tilde{S}_B W)} + \lambda \|W\|_M \\ \text{s. t. } W^T W &\in \Psi \end{aligned} \quad (6)$$

而对于 MSSP_ML, \tilde{S}_W 和 \tilde{S}_B 以 S_W 和 S_B 代替。

3.3 参数分析

在选择超参数的时候, 如果没有足够的经验和专家来帮助, 通常没有很好的方法, 只能使用交叉验证的方法, 这使得模型训练的代价很大。对于 MSSP_LDA 和 MSSP_ML, 根据损失函数的变化趋势和 $\|W\|_M$ 的取值范围推测 λ 随着 d 的变化趋势。

假设总能找到最佳的 d 个投影方向, 则损失函数的最小值随着 d 的增加而增加。这意味着损失函数的下界随着 d 的增加而增加。为了保证融合特征的判别性, 损失函数必然有一个上界。因此, 在保证判别性的前提下, 损失函数的取值范围随着 d 的增加而减小。

对于 $\|W\|_M$, 可以得到:

$$\|W\|_M = \sum_{i=1}^p \|W_i\|_{2,1} = \sum_{i=1}^p \sum_{j=1}^d \|W_i^j\|_2 = \sum_{j=1}^d \sum_{i=1}^p \|W_i^j\|_2 \quad (7)$$

由于 W 是列正交的, 故有:

$$\sum_{i=1}^p \|W_i^j\|_2^2 = \|W^j\|_2^2 = 1 \quad (8)$$

很容易得到:

$$1 \leq \sum_{i=1}^p \|W_i^j\|_2 \leq \sqrt{p} \Rightarrow d \leq \|W\|_M \leq d\sqrt{p} \quad (9)$$

$\|W\|_M$ 的取值范围随着 d 的增大而迅速增大, 这与损失函数取值范围的变化趋势是相反的。这意味着随着 d 的增大, $\|W\|_M$ 对 $f(W)$ 的影响很可能比损失函数要大。因此, 可以推断, 为了确保损失函数和正则化之间的最佳平衡, 最佳的超参数会随着 d 的增加而减小, 而实验结果也验证了上述推测。

3.4 优化求解

采用在 Stiefel 流形上的梯度下降法得到局部最优解, MSSP 框架优化算法简介如下:

Input: Views $X_1 \cdots X_v$, hyper-parameter λ

Step1: Choose loss function of MSSP_ML (S_W and S_B) or

MSSP_LDA (\tilde{S}_W and \tilde{S}_B)

Step2: Gradient descent over Stiefel manifold

(a) Initialize W

(b) loop

Calculate ∇ according to (10)

Calculate ∇_T according to (13)

Select β with Armijo rule.

Calculate \tilde{W} according to (14)

Calculate W_{new} according to (15)

end loop until converge

Output: Projection matrices $W_1 \cdots W_v$

梯度可以计算如下:

$$\nabla = \frac{\partial f}{\partial W} = \frac{2(T_2 S_W - T_1 S_B)}{T_2^2} + \lambda \begin{bmatrix} W_1 D_1 \\ \vdots \\ W_v D_v \end{bmatrix} \quad (10)$$

$$\begin{cases} T_1 = \text{tr}(W^T S_W W) \\ T_2 = \text{tr}(W^T S_B W) \end{cases} \quad (11)$$

$$D_i = \text{diag}\left(\frac{1}{\|W_i^1\|}, \cdots, \frac{1}{\|W_i^d\|}\right) \quad (12)$$

已经介绍了切空间 T_M , 它是在点 M 上沿着流形变化的所有方向的集合。因此 W 实际不是沿着 ∇ 变化。根据文献[21-22], 将梯度 ∇ 映射到切空间 T_W :

$$\nabla_T = \nabla - W \nabla^T W^T \quad (13)$$

然后使用 ∇_T 更新 W 得到:

$$\tilde{W} = W - \beta \nabla_T \quad (14)$$

其中, β 是步长, 通过 Armijo 规则自适应选择。

对任意非零步长 β , \tilde{W} 将跳出 Stiefel 流形, 因此需要

一个回拉来映射到流形上。根据文献[23], 应用奇异值分解(singular value decomposition, SVD)得到:

$$\tilde{W} = [U_1 \ U_2] \Lambda V^T, W_{\text{new}} = U_1 V^T$$

$$U_1 \in R^{r \times d}, U_2 \in R^{r \times (r-d)} \quad (15)$$

4 实验

4.1 实验设置

在实验中, 将 MSSP_LDA 和 MSSP_ML 与以下几种算法进行比较。(1) 单视图的 LDA(S_LDA): 在每个视图以及所有视图的拼接视图上运行 LDA, 报告最好的结果。(2) EMRSLRA: 其超参数选择的方法采用文献[6]的方式。(3) View-Consistency MVDA($VcM-VDA$)^[24]: 在 MVDA 的基础上, 约束所有的投影矩阵互相接近。(4) SDMCCA^[25]: 一种 MCCA 的监督型改进算法。

算法的性能是通过 KNN 的精确度来评估的, 其中 K 用 $c+1$ 。每个实验随机选取 80% 的样本进行训练, 其余 20% 用于测试。每个实验进行 10 次, 结果取平均值, 通过交叉验证选择超参数。

4.2 数据集

表 2 介绍了五个真实数据集: NUS-WIDE-LIFE^[26]、MFD^[27]、AWA^[28]、ADNI (sdin. loni. usc. edu)、WebKB^[29]。NUS-WIDE-LIFE 数据集是 NUS-WIDE 数据集的一个子集, 由 28 807 个训练图像以及 28 808 个测试图像组成, 每个图像有 81 维的标签向量。实验中只使用物体图像, 由于处理的是单标签图像, 所以进一步去掉了零标签或多个标签的图像。在丢弃稀有图像的类别之后, 有以下九类: 鸟、船、花、岩石、太阳、塔、玩具、树和车辆, 其中 10 600 个图像用于训练, 7 094 个图像用于测试。对 AWA 数据集, 选择了羚羊、灰熊、虎鲸、河狸和达尔马提亚狗五种动物的数据。从每种动物数据中随机选择不多于 200 个样本, 为避免过拟合用 PCA 将数据降到 50 维。对 ADNI 数据集, 选择患病者和正常人两类数据进行实验, MRI、FDG-PET 和 AV45-PET 数据按文献[30]方式进行预处理。对 WebKB 数据集, 先用 PCA 降到 50 维。对 MFD 数据集, 去掉维数较少的 mor 和 zer 两个视图。

表 2 数据集

数据集	视图/维数	类数/样本数
MFD	fou/76 fac/216 kar/64 pix/240 zer/47 mor/6	10/200×10
AWA	cq/2688 lss/2000 phog/252 rgift/2000 sift/2000 surf/2000	5/200×5
ADNI	VBM/116 FDG/116 AV/116	2/211, 160
NUS-WIDE-LIFE	ch/65 b-wcm/226 ca-c/145 edh/74 wt/129	9/10 600+7 094
WebKB	text/3000 link/1840	2/230, 821

4.3 实验结果与分析

表 3 中的实验结果表明:

(1) 基于 MSSP 框架的两种方法,分类的准确率显著好于其他算法,说明同时关注一致性和互补性对分类效果是有促进作用的;

(2) MSSP_LDA 在四数据集上结果显著好于 V_c -MVDA,同样是 LDA 在多视图上的推广,基于 MSSP 的框架具有更好的性能。一方面是由于 MSSP 的框架结构简单,LDA 在推广的过程中形式和原理都没有发

生变化,另一方面结构化稀疏正则化使得推广更灵活,可以适应多视图数据的结构特点,可以同时提取共享信息和独有信息中对于分类效果有促进作用的信息;

(3) 在 ADNI 数据集上 MSSP 两种算法效果远高于其他几种算法,这是因为数据本身对稀疏性有偏好,稀疏的方法在这种数据上通常有较好的效果,而 MSSP 的稀疏正则化项也能满足这种偏好,这是 MSSP 的形式本身的特点。

表 3 对比实验结果(均值 \pm 标准差)

Data	S_LDA	EMRSLRA	V_c MVDA	SDMCCA	MSSP_LDA	MSSP_ML
AWA	0.624 5 \pm 0.050 0	0.539 1 \pm 0.029 8	0.483 8 \pm 0.015 9	0.629 6 \pm 0.041 3	0.435 7 \pm 0.038 7	0.636 1 \pm 0.044 0
NUS	0.434 1 \pm 0.006 0	0.465 9 \pm 0.009 7	0.338 1 \pm 0.003 4	0.232 4 \pm 0.009 0	0.396 7 \pm 0.015 4	0.476 7 \pm 0.007 3
WebKB	0.975 2 \pm 0.010 5	0.983 8 \pm 0.058 6	0.958 6 \pm 0.008 9	0.986 2 \pm 0.010 3	0.986 6 \pm 0.007 7	0.990 5 \pm 0.007 2
MFD	0.955 3 \pm 0.003 7	0.958 0 \pm 0.010 5	0.865 4 \pm 0.013 6	0.951 8 \pm 0.008 1	0.981 5 \pm 0.029 8	0.972 5 \pm 0.001 9
ADNI	0.779 7 \pm 0.060 4	0.771 6 \pm 0.044 5	0.762 2 \pm 0.028 8	0.805 4 \pm 0.049 6	0.890 5 \pm 0.040 9	0.944 6 \pm 0.024 9

4.4 参数分析

图 2 和图 3 分别为 MSSP_ML 在 ANDI 和 AWA 数据集上,最佳超参数随投影维数的变化趋势。可以发现,最佳的超参数会随着 d 的增加而减小,并且随着 d 的增加,减小的速度变快,这验证了前文中对超参数的分析和推测。

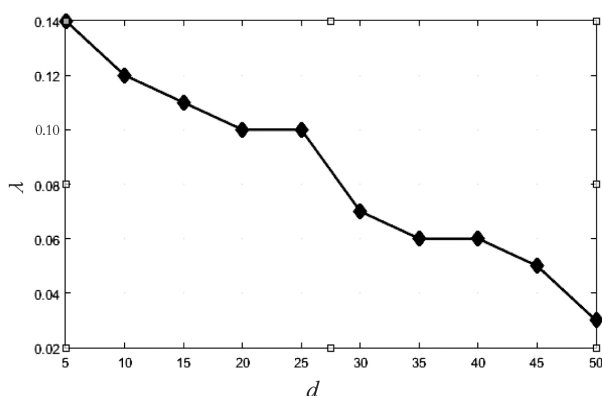


图 2 MSSP_ML 在 ANDI 上最佳超参数变化趋势

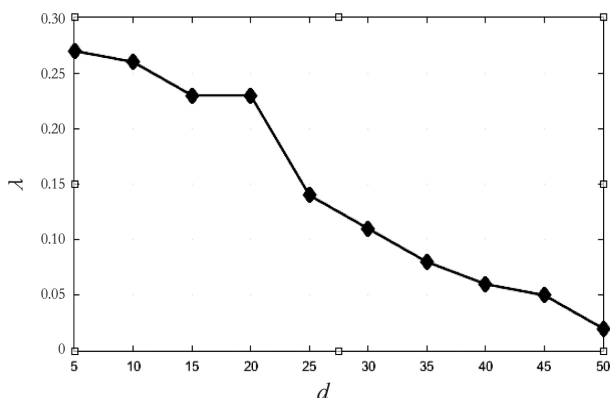


图 3 MSSP_ML 在 AWA 上最佳超参数变化趋势

5 结束语

根据信息保持的两个假设和一致性与互补性两个原则,将多视图降维方法细分为六类,提出了一个新的多重结构化稀疏投影框架 MSSP,并分析了超参数的变化趋势。在真实数据集上的对比实验,验证了 MSSP 的有效性和对超参数变化趋势的推测。在未来的工作中,可以尝试将更多的单视图方法推广到 MSSP 框架中。

参考文献:

- [1] KUMAR A, RAI P. Co-regularized multi-view spectral clustering [C]//International conference on neural information processing systems. Granada, Spain: Curran Associates Inc. 2011: 1413-1421.
- [2] GUO Yuhong. Convex subspace representation learning from multi-view data [C]//Twenty-seventh AAAI conference on artificial intelligence. Bellevue, Washington: AAAI Press, 2013: 387-393.
- [3] 孙权森, 曾生根, 王平安, 等. 典型相关分析的理论及其在特征融合中的应用 [J]. 计算机学报, 2005, 28(9): 1524-1533.
- [4] SUN Tingkai, CHEN Songcan, YANG Jingyu, et al. A supervised combined feature extraction method for recognition [C]//Proceedings of the IEEE international conference on data mining. Pisa, Italy: IEEE, 2008: 1043-1048.
- [5] KAN Meina, SHAN Shiguang, ZHANG Haihong, et al. Multiview discriminant analysis [C]//European conference on computer vision. Berlin: Springer, 2012: 808-821.
- [6] ZHANG Lefei, ZHANG Qian, ZHANG Liangpei, et al. En-

- semble manifold regularized sparse low-rank approximation for multiview feature embedding [J]. *Pattern Recognition*, 2015, 48(10): 3102–3112.
- [7] JIA Y, SALZMANN M, DARRELL T. Factorized latent spaces with structured sparsity [C]//Conference on neural information processing systems 2010. Vancouver, British Columbia, Canada [s. n.]: 2010: 982–990.
- [8] LIU Jing, JIANG Yu, LI Zechao, et al. Partially shared latent factor learning with multiview data [J]. *IEEE Transactions on Neural Networks & Learning Systems*, 2015, 26(6): 1233–1246.
- [9] 张荣, 孙权森. 基于核稀疏保持投影的典型相关分析算法[J]. *数据采集与处理*, 2017, 32(1): 111–118.
- [10] RUPNIK J, SHAWA-TAYLOR J. Multi-view canonical correlation analysis [C]//Conference on data mining and data warehouses. Taylor [s. n.]: 2010: 1–4.
- [11] EISENSTEIN J, O’CONNOR B, SMITH N A, et al. A latent variable model for geographic lexical variation [C]//Conference on empirical methods in natural language processing. Cambridge, Massachusetts: Association for Computational Linguistics, 2010: 1277–1287.
- [12] GUO Yiwen, DING Xiaoqing, LIU Changsong, et al. Sufficient canonical correlation analysis [J]. *IEEE Transactions on Image Processing*, 2016, 25(6): 2610–2619.
- [13] 陈理国, 刘超. 基于高斯过程的缺陷定位方法[J]. *软件学报*, 2014, 25(6): 1169–1179.
- [14] SHON A, GROCHOW K, HERTZMANN A, et al. Learning shared latent structure for image synthesis and robotic imitation [C]//Advances in neural information processing systems. Vancouver, British Columbia, Canada [s. n.]: 2006: 1233–1240.
- [15] 成亚玲, 谭爱平, 张敏. 混合多距离图像的线性判别分析人脸识别算法[J]. *系统仿真学报*, 2016, 28(9): 2254–2259.
- [16] ZADEH P H, HOSSEINI R, SRA S. Geometric mean metric learning [C]//International conference on machine learning. [s. l.]: [s. n.]: 2016: 2464–2471.
- [17] SUN Quansen, ZENG Shenggen, LIU Yan, et al. A new method of feature fusion and its application in image recognition [J]. *Pattern Recognition*, 2005, 38(12): 2437–2448.
- [18] 陶卿, 高乾坤, 姜纪远, 等. 稀疏学习优化问题的求解综述[J]. *软件学报*, 2013, 24(11): 2498–2507.
- [19] 傅俊鹏, 陈秀宏, 葛晓倩. L2, L1 范数正则化的广义核判别分析及其人脸识别[J]. *计算机科学与探索*, 2017, 11(1): 124–133.
- [20] CUNNINGHAM J P, GHAHRAMANI Z. Linear dimensionality reduction: survey, insights and generalizations [J]. *Journal of Machine Learning Research*, 2015, 16: 2859–2900.
- [21] EDELMAN A, ARIAS T A, SMITH S T. The geometry of algorithms with orthogonality constraints [J]. *SIAM Journal on Matrix Analysis and Applications*, 1998, 20(2): 303–353.
- [22] 吴秋峰, 刘振忠. Stiefel 流形上的梯度下降法[J]. *应用数学学报*, 2012, 35(4): 719–727.
- [23] KANEKO T, FIORI S, TANAKA T. Empirical arithmetic averaging over the compact stiefel manifold [J]. *IEEE Transactions on Signal Processing*, 2013, 61(4): 883–894.
- [24] KAN M, SHAN S, ZHANG H, et al. Multi-view discriminant analysis [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(1): 188–194.
- [25] YUAN Yunhao, LI Yun, JI Hongkun, et al. Semi-discriminative multiview canonical correlation analysis for recognition [M]//Intelligent data engineering and automated learning. [s. l.]: Springer International Publishing, 2016.
- [26] CHUA T S, TANG J, HONG R, et al. NUS-WIDE: a real-world web image database from National University of Singapore [C]//ACM international conference on image and video retrieval. Santorini, Fira, Greece: ACM, 2009: 48.
- [27] VAN BREUKELEN M, DUIN R P W, TAX D M J, et al. Handwritten digit recognition by combined classifiers [J]. *Kybernetika*, 1998, 34(4): 381–386.
- [28] LAMPERT C H, NICKISCH H, HARMELING S. Learning to detect unseen object classes by between-class attribute transfer [C]//IEEE conference on computer vision and pattern recognition. Miami, FL, USA: IEEE, 2009: 951–958.
- [29] CHEN X, YIN H, JIANG F, et al. Multi-view dimensionality reduction based on universum learning [J]. *Neurocomputing*, 2018, 275: 2279–2286.
- [30] HAO X, YAO X, YAN J, et al. Identifying multimodal intermediate phenotypes between genetic risk factors and disease status in alzheimer’s disease [J]. *Neuroinformatics*, 2016, 14(4): 439–452.