

## 基于增强稀疏性特征选择的网络图像标注\*

史彩娟<sup>1,2</sup>, 阮秋琦<sup>1</sup>

<sup>1</sup>(北京交通大学 信息科学研究所, 北京 100044)

<sup>2</sup>(华北理工大学 信息工程学院, 河北 唐山 063009)

通讯作者: 史彩娟, E-mail: shicaijuan2011@gmail.com, http://www.bjtu.edu.cn

**摘要:** 面对网络图像的爆炸性增长, 网络图像标注成为近年来一个热点研究内容, 稀疏特征选择在提升网络图像标注效率和性能方面发挥着重要的作用. 提出了一种增强稀疏性特征选择算法, 即, 基于  $l_{2,1/2}$  矩阵范数和共享子空间的半监督稀疏特征选择算法 (semi-supervised sparse feature selection based on  $l_{2,1/2}$ -matrix norm with shared subspace learning, 简称 SFSL) 进行网络图像标注. 在 SFSL 算法中, 应用  $l_{2,1/2}$  矩阵范数来选取最稀疏和最具判别性的特征, 通过共享子空间学习, 考虑不同特征之间的关联信息. 另外, 基于图拉普拉斯的半监督学习, 使 SFSL 算法同时利用了有标签数据和无标签数据. 设计了一种有效的迭代算法来最优化目标函数. SFSL 算法与其他稀疏特征选择算法在两个大规模网络图像数据库上进行了比较, 结果表明, SFSL 算法更适合于大规模网络图像的标注.

**关键词:** 网络图像标注; 稀疏特征选择;  $l_{2,1/2}$  矩阵范数; 共享子空间学习; 半监督学习

**中图法分类号:** TP391

中文引用格式: 史彩娟, 阮秋琦. 基于增强稀疏性特征选择的网络图像标注. 软件学报, 2015, 26(7): 1800-1811. <http://www.jos.org.cn/1000-9825/4687.htm>

英文引用格式: Shi CJ, Ruan QQ. Feature selection with enhanced sparsity for Web image annotation. Ruan Jian Xue Bao/ Journal of Software, 2015, 26(7): 1800-1811 (in Chinese). <http://www.jos.org.cn/1000-9825/4687.htm>

## Feature Selection with Enhanced Sparsity for Web Image Annotation

SHI Cai-Juan<sup>1,2</sup>, RUAN Qiu-Qi<sup>1</sup>

<sup>1</sup>(Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China)

<sup>2</sup>(College of Information Engineering, North China University of Science and Technology, Tangshan 063009, China)

**Abstract:** In dealing with the explosive growth of web images, Web image annotation has become a critical research issue in recent years. Sparse feature selection plays an important role in improving the efficiency and performance of Web image annotation. In this paper, a feature selection framework is proposed with enhanced sparsity for Web image annotation. The new framework, termed as semi-supervised sparse feature selection based on  $l_{2,1/2}$ -matrix norm with shared subspace learning (SFSL), selects the most sparse and discriminative features by utilizing  $l_{2,1/2}$ -matrix norm and obtains the correlation between different features via shared subspace learning. In addition, SFSL uses graph Laplacian semi-supervised learning to exploit both labeled and unlabeled data simultaneously. An efficient iterative algorithm is designed to optimize the objective function. SFSL method is compared to other feature selection algorithms on two Web image datasets and the results indicate it is suitable for large-scale Web image annotation.

**Key words:** Web image annotation; sparse feature selection;  $l_{2,1/2}$ -matrix norm; shared subspace learning; semi-supervised learning

面对网络图像的爆炸性增长, 网络图像标注成为近年来一个研究热点. 通常, 图像被表示为各种不同类型的特征, 因此, 特征选择成为提高网络图像标注性能的一个重要手段. 最近, 稀疏特征选择以其较好的特征选择

\* 基金项目: 国家重点基础研究发展计划(973)(2012CB316304); 国家自然科学基金(61471032, 61472030); 河北省高等学校科学技术研究青年基金(QN2014026); 教育部创新团队项目(IRT201206); 北京市现代信息科学与网络技术重点实验室资助项目

收稿时间: 2013-08-11; 定稿时间: 2014-07-09

性能和较低的计算复杂度被广泛用于网络图像标注<sup>[1-4]</sup>中.稀疏特征选择就是采用各种稀疏模型来实现特征选择和数据的稀疏表示.

最著名的稀疏模型是  $l_1$  范数模型(lasso)<sup>[5]</sup>,蔡等人<sup>[6]</sup>采用  $l_1$  正则化回归模型进行特征选择.尽管  $l_1$  范数计算方便,但是所选取的特征稀疏性不强,因此, $l_1$  范数模型被扩展到了  $l_p(0 < p < 1)$  范数模型, $l_p$  范数模型比  $l_1$  范数模型具有更好的稀疏性.最近,徐等人<sup>[7,8]</sup>指出:当  $p=1/2$  时, $l_p$  范数模型,也即  $l_{1/2}$  范数模型具有最好的稀疏性.赵等人<sup>[9]</sup>将  $l_{1/2}$  范数模型用于逻辑回归的处理上.然而,上述模型都是对特征进行逐一的选取,忽略了特征之间的关联信息.聂等人<sup>[10]</sup>提出  $l_{2,1}$  范数模型进行特征选择,从所有的数据点中联合地选取特征;马等人<sup>[4]</sup>将  $l_{2,1}$  范数应用到了图像标注.但是  $l_{2,1}$  范数模型是基于  $l_1$  范数模型的,因此所选取的特征还不够稀疏.

通常,一幅图像包含几个语义概念,因此图像标注实质上是一个多标签分类问题.Ando 等人<sup>[11]</sup>提出不同的标签之间存在共享子空间,该共享子空间可以利用标签关联提升多标签分类的性能.杨等人<sup>[12]</sup>将共享子空间学习应用到多标签图像标注.同理,不同特征之间的关联信息也可以通过共享子空间学习来获得.马等人<sup>[13]</sup>将共享子空间学习应用到了稀疏特征选择,进而提升了网络图像标注的性能.

监督学习方法中,人工标注训练数据既费时又费力,因此不适合大规模网络图像标注.无标签训练数据易于获取,同时能够促进标注性能的提高.为了减少标注训练数据的人力投入成本,同时充分利用未标注训练数据,许多半监督学习方法被广泛应用到网络图像标注<sup>[12,14,15]</sup>中.在半监督学习方法中,最常用的是基于图拉普拉斯半监督学习.由于简单方便,KNN 图拉普拉斯半监督学习被广泛采用.马等人<sup>[4]</sup>提出了一种半监督特征选择方法进行图像标注;杨等人<sup>[12]</sup>将共享结构学习与基于图的半监督学习相结合进行网络图像的标注;唐等人<sup>[14]</sup>提出基于 KNN 稀疏图的半监督学习方法进行有噪声标签网络图像标注.

由于  $l_{1/2}$  范数模型比  $l_1$  范数模型选取的特征更为稀疏有效<sup>[7,8]</sup>,我们将  $l_{2,1}$  范数扩展到  $l_{2,1/2}$  矩阵范数,利用该模型进行更加有效的稀疏特征选择.考虑到不同特征之间的关联信息,在特征选择过程中加入共享子空间学习来提升特征选择的效果.面对大量的网络图像以及无标签训练图像,半监督特征选择方法更加有效.因此,本文结合以上 3 点内容,提出一种增强稀疏性特征选择方法进行网络图像标注,即,基于  $l_{2,1/2}$  矩阵范数和共享子空间的半监督稀疏特征选择算法(semi-supervised sparse feature selection based on  $l_{2,1/2}$ -matrix norm with shared subspace learning,简称 SFSL).该算法在两个大规模网络图像数据库 NUS-WIDE<sup>[16]</sup>和 MSRA-MM 2.0<sup>[17]</sup>上进行了实验,结果表明:SFSL 算法优于现有的其他稀疏特征选择算法,适合大规模网络图像标注.

## 1 相关工作

本节介绍与本文相关的 3 个主要工作,即:稀疏特征选择、共享子空间学习和图拉普拉斯半监督学习.

### 1.1 稀疏特征选择

稀疏特征选择就是采用各种不同的稀疏模型来选取最具判别性的特征,从而得到数据的稀疏表示.稀疏特征选择不仅可以提升网络图像标注性能,同时还可以降低计算的复杂度.

最著名的稀疏模型是  $l_1$  范数,该模型是凸问题,易于求解,但是所选取的特征不够稀疏.最近,徐等人指出, $l_{1/2}$  范数模型比  $l_1$  范数具有更好的稀疏性<sup>[7,8]</sup>.然而, $l_{1/2}$  范数模型和  $l_1$  范数模型在稀疏特征选择过程中忽略了特征之间的关联信息.聂等人<sup>[10]</sup>提出  $l_{2,1}$  范数,该模型从所有数据点中进行特征选择;王等人<sup>[18]</sup>将  $l_{2,1}$  范数扩展为  $l_{2,p}$  范数( $0 < p \leq 1$ ).

设  $G \in R^{d \times c}$  为稀疏特征选择映射矩阵,其  $l_{2,p}$  矩阵范数定义为

$$\|G\|_{2,p} = \left( \sum_{i=1}^d \|g^i\|_p^p \right)^{1/p}, p \in (0,1] \quad (1)$$

很显然,当  $p=1$  时, $l_{2,p}$  矩阵范数即为  $l_{2,1}$  矩阵范数.注意: $l_{2,p}$  矩阵范数是伪范数,不满足三角不等式.另外, $l_{2,p}$  矩阵范数是非凸的,不能直接采用文献[10]中的方法进行求解.

当  $p=1/2$  时, $l_{2,p}$  矩阵范数即为  $l_{2,1/2}$  矩阵范数  $\|G\|_{2,1/2}$ . 本文将  $l_{2,1/2}$  矩阵范数应用到 SFSL 算法中.

## 1.2 共享子空间学习

在多标签学习问题中,Ando 等人<sup>[11]</sup>提出:在原始的特征空间中存在一个共享子空间,该共享子空间与原始特征空间可以同时同时对图像语义概念进行预测.

给定特征向量  $x \in R^d$  和预测函数  $f$ ,则共享特征子空间学习可以定义为下式:

$$f(x) = v^T x + p^T U^T x \quad (2)$$

上式中,  $v \in R^d$  和  $p \in R^r$  是权重向量,  $U \in R^{d \times r}$  是所有特征的共享子空间.

为了获得预测函数  $f$ ,目标函数可写为

$$\min_f \sum_{i=1}^n \text{loss}(f(x_i), y_i) + \gamma R(f) \quad (3)$$

上式中,  $\text{loss}()$  是误差函数,  $\gamma R(f)$  是正则项.

假设多标签学习中的图像包含  $c$  个概念,第  $t$  个概念包含  $h_t$  个训练数据  $\{x_i\}_{i=1}^{h_t}$ ,对应的标签为  $\{y_i\}_{i=1}^{h_t}$ .将公式(2)带入公式(3),则此时的目标函数为

$$\min_{\{v_t, p_t, U\}} \sum_{t=1}^c \left( \frac{1}{h_t} \sum_{i=1}^{h_t} \text{loss}((v_t + U p_t)^T x_i, y_i) + \gamma R(\{v_t, p_t\}) \right) \quad (4)$$

s.t.  $U^T U = I$

公式(4)中,  $U^T U = I$  为限定项,便于问题的求解.

由于共享子空间学习在多标签数据中的有效性,本文将共享子空间学习应用到特征选择框架 SFSL 中.

## 1.3 图拉普拉斯半监督学习

半监督学习可以同时利用有标签数据和无标签数据,因此非常适合仅有很少标签数据的网络图像标注.一方面,半监督学习可以节省标注大量图像的人力,同时可以利用无标签数据来提升网络图像标注的性能.基于图拉普拉斯半监督学习得到了广泛研究,其中, KNN 图及相关方法被广泛用于网络图像标注<sup>[4,12,14]</sup>中.

设  $X = [x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_n]^T$  为训练图像的特征矩阵,共有  $n$  个训练数据,其中包含  $m$  个有标签数据,  $x_i \in R^d$  ( $1 \leq i \leq n$ ) 表示第  $i$  个图像的特征向量.设  $Y = [y_1, y_2, \dots, y_m, y_{m+1}, \dots, y_n]^T$  为训练图像的标签矩阵,  $c$  表示类别数,  $y_i \in R^c$  ( $1 \leq i \leq n$ ) 表示第  $i$  个标签向量.设  $Y_{ij}$  表示  $y_i$  的第  $j$  个数据,如果  $x_i$  是第  $j$  类,则  $Y_{ij} = 1$ ; 否则,  $Y_{ij} = 0$ .如果  $x_i$  是无标签数据,即  $i > m$ ,则  $y_i$  为全 0 向量.

定义图模型  $S$ ,其元素  $S_{ij}$  表示  $x_i$  与  $x_j$  之间的相似性:

$$S_{ij} = \begin{cases} 1, & x_i \text{ 和 } x_j \text{ 是 } k \text{ 近邻} \\ 0, & \text{其他} \end{cases} \quad (5)$$

设  $D_{ii} = \sum_{j=1}^n S_{ij}$  为对角矩阵,则图拉普拉斯矩阵可以表示为  $L = D - S$ .

朱等人<sup>[19,20]</sup>提出了同时利用有标签图像和无标签图像的半监督直推式分类算法.设所有训练数据的标签预测矩阵为  $F = [f_1, f_2, \dots, f_n]^T \in R^{n \times c}$ ,其中  $f_i \in R^c$  ( $1 \leq i \leq n$ ) 为  $x_i$  的预测标签.聂等人<sup>[21]</sup>指出:  $F$  应该与训练数据的真实标签保持一致,同时,在图模型上具有光滑性.因此,可以通过最小化下面的目标函数来获得预测标签矩阵  $F$ .

$$\arg \min_F \text{Tr}(F^T L F) + \text{Tr}((F - Y)^T Q (F - Y)) \quad (6)$$

上式中,对角矩阵  $Q \in R^{n \times n}$  为决策规则矩阵.如果  $x_i$  为有标签数据,则  $Q$  的对角元素  $Q_{ii} = \infty$ ; 否则,其对角元素  $Q_{ii} = 1$ .决策规则矩阵  $Q$  使预测标签矩阵  $F$  与训练数据的真实标签矩阵  $Y$  保持一致.

基于图的直推式算法被广泛应用于多媒体领域<sup>[22-24]</sup>.与以前的半监督学习算法不同<sup>[25,26]</sup>,这里,我们根据标签预测矩阵来最小化预测误差,并将其应用到本文提出的 SFSL 模型中.

## 2 基于 $l_{2,1/2}$ 矩阵范数和共享子空间的半监督稀疏特征选择算法

本节中,我们提出一种增强稀疏性特征选择算法,即,基于  $l_{2,1/2}$  矩阵范数和共享子空间的半监督稀疏特征选

择算法,简称 SFSLs. 首先提出该算法,然后给出了算法的求解过程及算法的收敛性分析.

## 2.1 算法提出

算法的基本思路:将  $l_{2,1/2}$  矩阵范数、共享子空间学习和基于图拉普拉斯半监督学习相结合,得到稀疏特征选择模型 SFSLs.

给定  $X=[x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_n]^T$  为训练图像的特征矩阵,  $m$  表示有标签数据的数目,  $n$  为所有训练数据数目,  $x_i \in R^d (1 \leq i \leq n)$  表示第  $i$  个图像的特征向量. 设  $Y=[y_1, y_2, \dots, y_m, y_{m+1}, \dots, y_n]^T$  为训练图像的标签矩阵,  $c$  表示类别数,  $y_i \in R^c (1 \leq i \leq n)$  表示第  $i$  个图像的标签向量. 设  $Y_{ij}$  表示  $y_i$  的第  $j$  个数据, 如果  $x_i$  是第  $j$  类, 则  $Y_{ij}=1$ ; 否则,  $Y_{ij}=0$ . 如果  $x_i$  是无标签数据, 即  $i > m$ , 则  $y_i$  为全 0 向量.

定义  $V=[v_1, v_2, \dots, v_c] \in R^{d \times c}$ ,  $P=[p_1, p_2, \dots, p_c] \in R^{r \times c}$  和  $U \in R^{d \times r}$ , 这里,  $r$  表示共享子空间的维度.

定义映射矩阵为  $G=V+UP$ , 其中,  $G \in R^{d \times c}$ , 则公式(4)可以写为

$$\begin{aligned} \min_{V, P, U} \text{loss}(G^T X, Y) + \gamma R(V, P) \\ \text{s.t. } U^T U = I \end{aligned} \quad (7)$$

将  $l_{2,1/2}$  矩阵范数  $\|G\|_{2,1/2}^{1/2}$ 、共享特征子空间学习(7)和基于图拉普拉斯半监督学习(6)相结合,同时采用最小均方误差函数作为损失函数,则得到稀疏特征选择模型 SFSLs 的目标函数为

$$\begin{aligned} \arg \min_{F, G, U, P} \text{tr} F^T L F + \text{tr} (F - Y)^T Q (F - Y) + \alpha [\|X^T G - F\|_F^2 + \lambda \|G\|_{2,1/2}^{1/2} + \mu \|G - UP\|_F^2] \\ \text{s.t. } U^T U = I \end{aligned} \quad (8)$$

上式中,  $\lambda \|G\|_{2,1/2}^{1/2} + \mu \|G - UP\|_F^2$  为正则项, 其中,  $\|G\|_{2,1/2}^{1/2}$  保证 SFSLs 模型选取的特征更加稀疏, 更具判别性;  $\|G - UP\|_F^2$  共享子空间保证 SFSLs 模型在特征选择过程中充分考虑了不同特征之间的相关性.

值得注意的是, 本文提出的算法与文献[12]的算法和文献[4]的算法不同. 文献[12]中, 算法不包含稀疏特征选择, 我们提出的算法采用  $l_{2,1/2}$  矩阵范数稀疏模型进行稀疏特征选择. 文献[4]中介绍的算法虽然包含稀疏选择过程, 但是采用的稀疏模型为  $l_{2,1}$  矩阵范数, 稀疏性较差, 同时没有充分考虑不同特征之间的相关性. 由于  $l_{2,1/2}$  矩阵范数是非凸的, 因此不能直接采用文献[12]和文献[4]中的方法来求解. 下面我们提出了一种有效的迭代方法来求解 SFSLs 的目标函数.

## 2.2 优化求解

由于目标函数(8)中包含  $l_{2,1/2}$  矩阵范数, 因此该目标函数是非凸的, 我们采用如下方法来求解公式(8).

给定  $G=[g^1, \dots, g^d]^T$ , 设  $D$  为对角矩阵, 其对角元素为  $D_{ii} = \frac{1}{4} \|g^i\|_2^{3/2}$ , 则:

$$\|G\|_{2,p}^p = \frac{2}{p} \text{Tr}(G^T D G), \|G\|_{2,1/2}^{1/2} = 4 \text{Tr}(G^T D G).$$

对于任意矩阵  $A$ , 有:  $\|A\|_F^2 = \text{Tr}(A^T A)$ .

将目标函数(8)中的矩阵范数转换为迹操作, 则目标函数等价于:

$$\left. \begin{aligned} \arg \min_{F, G, U, P} \text{Tr}(F^T L F) + \text{Tr}((F - Y)^T Q (F - Y)) + \\ \alpha [\text{Tr}((X^T G - F)^T (X^T G - F)) + 4\lambda \text{Tr}(G^T D G) + \mu \text{Tr}((G - UP)^T (G - UP))] \\ \text{s.t. } U^T U = I \end{aligned} \right\} \quad (9)$$

对公式(9)中的参数  $P$  求导, 并令等式为 0, 则有:

$$(U^T U P - U^T G) = 0 \Rightarrow P = U^T G \quad (10)$$

将公式(10)中的  $P$  带入到公式(9), 则目标函数为

$$\left. \begin{aligned}
 & \arg \min_{F,G,U} \text{Tr}(F^T L F) + \text{Tr}((F-Y)^T Q(F-Y)) + \\
 & \quad \alpha [\text{Tr}((X^T G - F)^T (X^T G - F)) + 4\lambda \text{Tr}(G^T D G) + \mu \text{Tr}((G - U U^T G)^T (G - U U^T G))] \Rightarrow \\
 & \arg \min_{F,G,U} \text{Tr}(F^T L F) + \text{Tr}((F-Y)^T Q(F-Y)) + \\
 & \quad \alpha [\text{Tr}((X^T G - F)^T (X^T G - F)) + 4\lambda \text{Tr}(G^T D G) + \mu \text{Tr}(G^T (I - U U^T)^T (I - U U^T) G)] \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (11)$$

因为  $(I - U U^T)(I - U U^T) = (I - U U^T)$ , 所以公式(11)可以写为

$$\left. \begin{aligned}
 & \arg \min_{F,G,U} \text{Tr}(F^T L F) + \text{Tr}((F-Y)^T Q(F-Y)) + \alpha [\text{Tr}((X^T G - F)^T (X^T G - F)) + \text{Tr}(G^T (4\lambda D + \mu I - \mu U U^T) G)] \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (12)$$

对公式(12)中的参数  $G$  求导,并令等式为 0,则有:

$$2 X X^T G - 2 X F + 2(4\lambda D + \mu I - \mu U U^T) G = 0 \Rightarrow (X X^T + 4\lambda D + \mu I - \mu U U^T) G = X F \Rightarrow G = (A - \mu U U^T)^{-1} X F \Rightarrow G = B^{-1} X F \quad (13)$$

上式中,  $A = X X^T + 4\lambda D + \mu I$ ,  $B = A - \mu U U^T$ ,  $B = B^T$ .

因此,公式(12)可重写为

$$\left. \begin{aligned}
 & \arg \min_{F,G,U} \text{Tr}(F^T L F) + \text{Tr}((F-Y)^T Q(F-Y)) + \alpha [\text{Tr}(G^T B G) - 2 \text{Tr}(G^T X F) + \text{Tr}(F^T F)] \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (14)$$

将公式(13)中的  $G$  带入到公式(14),则有:

$$\left. \begin{aligned}
 & \arg \min_{F,U} \text{Tr}(F^T L F) + \text{Tr}((F-Y)^T Q(F-Y)) + \alpha [\text{Tr}(F^T F) - \text{Tr}(F^T X^T B^{-1} X F)] \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (15)$$

对公式(15)中的  $F$  求导,并令等式为 0,则得到:

$$L F + Q F - Q Y + \alpha [F - X^T B^{-1} X F] = 0 \Rightarrow F = (L + Q + \alpha I - \alpha X^T B^{-1} X)^{-1} Q Y = (R - \alpha X^T B^{-1} X)^{-1} Q Y \quad (16)$$

其中,  $R = L + Q + \alpha I$ .

此时,目标函数可以转化为

$$\left. \begin{aligned}
 & \arg \max_U \text{Tr}(Y^T Q (R - \alpha X^T B^{-1} X)^{-1} Q Y) \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (17)$$

根据 Sherman-Woodbury-Morrison 定理<sup>[27]</sup>,有:

$$(R - \alpha X^T B^{-1} X)^{-1} = R^{-1} + \alpha R^{-1} X^T (B - \alpha X R^{-1} X^T)^{-1} X R^{-1}.$$

因此,公式(17)转化为

$$\left. \begin{aligned}
 & \arg \max_U \text{Tr}(Y^T Q R^{-1} X^T (B - \alpha X R^{-1} X^T)^{-1} X R^{-1} Q Y) = \arg \max_U \text{Tr}(Y^T Q R^{-1} X^T J^{-1} X R^{-1} Q Y) \\
 & \text{s.t. } U^T U = I
 \end{aligned} \right\} \quad (18)$$

上式中,  $J = B - \alpha X R^{-1} X^T$ .

$$J^{-1} = (B - \alpha X R^{-1} X^T)^{-1} = (A - \mu U U^T - \alpha X R^{-1} X^T)^{-1} = (C - \mu U U^T)^{-1} \quad (19)$$

其中,  $C = A - \alpha X R^{-1} X^T$ .

再次利用 Sherman-Woodbury-Morrison 定理,则有:

$$(C - \mu U U^T)^{-1} = C^{-1} + \mu C^{-1} U (I - \mu U^T C^{-1} U)^{-1} U^T C^{-1} = C^{-1} + \mu C^{-1} U (U^T (I - \mu C^{-1}) U)^{-1} U^T C^{-1} \quad (20)$$

很显然,  $C$  与  $U$  相互独立.因此,我们可以将目标函数(18)重写为

$$\left. \begin{aligned} \arg \max_U \operatorname{Tr}(Y^T Q R^{-1} X^T C^{-1} U (U^T (I - \mu C^{-1}) U)^{-1} U^T C^{-1} X R^{-1} Q Y) = \\ \arg \max_U \operatorname{Tr}(Y^T Q R^{-1} X^T C^{-1} U (U^T K U)^{-1} U^T C^{-1} X R^{-1} Q Y) \\ \text{s.t. } U^T U = I \end{aligned} \right\} \quad (21)$$

其中,  $K=I-\mu C^{-1}$ .

对于任意矩阵  $X, Y$  和  $Z$ , 有  $\operatorname{Tr}(XYZ)=\operatorname{Tr}(YZX)$ , 公式(21)转化为

$$\left. \begin{aligned} \arg \max_U \operatorname{Tr}((U^T K U)^{-1} U^T C^{-1} X R^{-1} Q Y Y^T Q R^{-1} X^T C^{-1} U) \Rightarrow \arg \max_U \operatorname{Tr}(U^T K U)^{-1} U^T H U \\ \text{s.t. } U^T U = I \end{aligned} \right\} \quad (22)$$

上式中,  $K=I-\mu C^{-1}, H=C^{-1} X R^{-1} Q Y Y^T Q R^{-1} X^T C^{-1}$ .

公式(22)可以转化为  $K^{-1}H$  的特征分解问题, 然而,  $U$  和  $G$  却不能直接从公式(22)求取. 因此, 在下面所示的算法 1 中提出了一种迭代算法来求取目标函数(22), 流程如下.

**算法 1.**

1. 输入训练图像特征矩阵  $X \in \mathbf{R}^{d \times n}$ , 对应的标签矩阵为  $Y \in \mathbf{R}^{m \times c}$ , 正则化参数  $\alpha, \lambda$  和  $\mu$ , 图拉普拉斯中的最近邻数  $k$ ;
2. 计算图拉普拉斯矩阵  $L \in \mathbf{R}^{n \times n}$ ;
3. 计算决策规则矩阵  $Q \in \mathbf{R}^{n \times n}$ ;
4. 计算  $R=L+Q+\alpha I$ ;
5. 设  $t=0$ , 初始化映射矩阵为随机矩阵  $G_0 \in \mathbf{R}^{d \times c}$ ;
6. 重复:

$$\text{计算对角矩阵 } D_t; D_t = \begin{bmatrix} \frac{1}{4 \|g_t^1\|_2^{3/2}} & & \\ & \dots & \\ & & \frac{1}{4 \|g_t^d\|_2^{3/2}} \end{bmatrix};$$

计算  $A_t=XX^T+4\lambda D_t+\mu I$ ;

计算  $C_t=A_t-\alpha X R^{-1} X^T$ ;

计算  $K_t=I-\mu C_t^{-1}$ ;

计算  $H_t=C_t^{-1} X R^{-1} Q Y Y^T Q R^{-1} X^T C_t^{-1}$ ;

对  $K_t^{-1}H_t$  进行特征分解;

计算  $B_t=A_t-\mu U_t U_t^T$ ;

计算  $F_t=(R-\alpha X^T B_t^{-1} X)^{-1} Q Y$ ;

更新  $G_{t+1}=B_t^{-1} X F_t$ ;

$t=t+1$ ;

7. 直到收敛, 返回最优映射矩阵  $G$ .

算法 1 中的迭代过程可以通过定理 1 证明其收敛于最优映射矩阵  $G$ , 为了证明定理 1, 需要引入引理 1. 关于引理 1 和定理 1 的证明可见附录.

**3 实验分析**

本节将所提算法 SFSL 在两个大规模图像数据库上进行了实验, 验证了其进行大规模网络图像标注的有效性.

### 3.1 数据库及评价标准

实验中使用两个网络图像数据库:NUS-WIDE<sup>[16]</sup>和 MSRA-MM2.0<sup>[17]</sup>.

NUS-WIDE 数据库是由新加坡国立大学媒体研究实验室于 2009 年创立的,该数据库包括 269 648 幅图像、81 个概念.实验中,我们选用 3 种类型特征,即:144 维颜色相关图、128 维小波纹理和 73 维边缘方向直方图.我们将这 3 种特征串接为 345 维特征来表示每一幅图像.

MSRA-MM2.0 数据库是由微软亚洲研究中心于 2009 年创立的,该数据库包括 50 000 幅图像、100 个概念.实验中,我们也选用 144 维颜色相关图、128 维小波纹理和 75 维边缘方向直方图这 3 种类型特征,并将其串接为 347 维特征来表示每一幅图像.

实验中,评估准则采用在多媒体领域里被广泛应用的平均准确率均值(MAP),其值越大,表示性能越好.

### 3.2 实验结果与分析

实验中随机采样 3 000 个图像形成训练数据集,剩余数据作为测试集.训练集中有标签数据数量为  $m$ ,分别为训练数据的 5%,10%,25%,50%和 100%,实验中分别记录相应的标注结果.随机进行 5 次实验,将 5 次结果的平均作为最终的标注结果.

将本文提出的 SFSLs 算法与现有的一些稀疏特征选择算法进行比较,这些算法包括贝叶斯  $l_1$  正则化稀疏多项式逻辑回归方法(SBMLR)<sup>[28]</sup>、联合  $l_{2,1}$  范数最小化特征选择方法(FSNM)<sup>[10]</sup>、稀疏性结构化特征选择方法(SFSS)<sup>[4]</sup>.实验中,方法 SFSLs 与 SFSS 可以同时实现特征选择与分类,而 FSNM 和 SBMLR 方法仅用于特征选择,因此特征选择后,通过最小均方回归模型进行分类.

#### 3.2.1 性能评估

本文所提算法 SFSLs 与其他算法进行网络图像标注的性能比较结果见表 1,在两个数据库上分别进行了实验,最好结果用粗体标识.

**Table 1** Performance comparison (MAP±standard deviation)

表 1 性能比较(MAP±标准偏差)

有标签数据 $n$	数据库	SFSLs	SFSS	FSNM	SMBLR
3000×5%	NUS-WIDE	<b>0.071±0.003</b>	0.060±0.002	0.054±0.002	0.047±0.002
	MSRA-MM	<b>0.046±0.002</b>	0.042±0.002	0.038±0.002	0.032±0.001
3000×10%	NUS-WIDE	<b>0.084±0.002</b>	0.077±0.002	0.070±0.002	0.058±0.002
	MSRA-MM	<b>0.054±0.001</b>	0.046±0.002	0.043±0.002	0.037±0.002
3000×25%	NUS-WIDE	<b>0.093±0.001</b>	0.088±0.001	0.081±0.001	0.071±0.001
	MSRA-MM	<b>0.061±0.001</b>	0.053±0.001	0.048±0.001	0.041±0.001
3000×50%	NUS-WIDE	<b>0.103±0.001</b>	0.100±0.002	0.096±0.001	0.091±0.002
	MSRA-MM	<b>0.068±0.001</b>	0.058±0.002	0.053±0.002	0.046±0.001
3000×100%	NUS-WIDE	<b>0.113±0.001</b>	0.108±0.001	0.106±0.001	0.102±0.003
	MSRA-MM	<b>0.075±0.001</b>	0.066±0.002	0.061±0.001	0.052±0.002

从表 1 我们可以得出以下结论:

- 首先,随着有标签数据量的增加,所有算法的性能都逐渐提高.例如,对于数据库 NUS-WIDE 上的 SFSLs 算法,当有标签数据占训练数据总量的 10%时,MAP 是 0.084;而当比例增加到 25%时,MAP 提高到 0.093;
- 其次,表 1 表明,本文算法 SFSLs 的性能优于其他算法.这说明 SFSLs 算法能够有效选取更具判别性的稀疏特征来提升图像标注性能.另外,也说明了该算法能够用较少的有标签训练数据的半监督学习方式来进行大规模网络图像标注.

对于 SFSLs 算法能够取得较好的图像标注性能,主要包括 3 个原因.

- (1) 采用  $l_{2,1/2}$  矩阵范数进行稀疏特征选择,使所选取的特征更加稀疏,更具判别性;
- (2) 算法中采用了共享子空间学习,充分利用了不同特征之间的关联信息;
- (3) 采用基于图拉普拉斯半监督学习,同时利用了有标签数据和无标签数据,无标签数据进一步提升了网

络图像标注的性能.

### 3.2.2 参数 $p$ 的影响

本文提出的 SFSLs 算法是基于  $l_{2,1/2}$  矩阵范数模型的,这里的  $l_{2,1/2}$  矩阵范数是  $l_{2,p}$  矩阵范数在  $p=1/2$  时的一个特例.本节专门设计实验来分析参数  $p$  取不同值时对标注性能的影响.由于  $p \in (0,1]$ ,我们选取典型的  $p$  值  $\{0.1, 0.25, 0.5, 0.75, 1\}$  在两个数据库上分别进行了实验.表 2 列出了采用不同  $p$  值时的标注结果,最好的结果用粗体字标识.

**Table 2** Performance comparison with different parameter  $p$  (MAP)

表 2 采用不同参数  $p$  的性能比较(MAP)

数据库	$p=0.1$	$p=0.25$	$p=0.5$	$p=0.75$	$p=1$
NUS-WIDE	0.081	0.083	<b>0.084</b>	0.079	0.082
MSRA-MM	0.051	0.052	<b>0.054</b>	0.051	0.053

实验中设定有标签数据量为训练数据总量的 10%.从表 2 我们得到以下结果:(1) 当  $p=1/2$  时,即本文算法 SFSLs,具有最好的标注性能;(2) 当  $p$  属于  $(0,0.5)$  时, $p$  越接近于 0,标注性能越差,这说明所选择的特征过于稀疏,丢失了有用信息;(3) 当  $p$  接近于 1 时,近似为 SFSS 算法.总之,该实验结果表明:本文所提算法 SFSLs(对应  $p=1/2$ )具有最好的稀疏选择特性,能够获得很好的标注性能.

### 3.2.3 无标签数据的影响

文本算法 SFSLs 中采用基于图拉普拉斯半监督学习,同时利用了有标签数据与无标签数据.本节设计了一个实验,验证无标签数据对标注性能的影响.为此,实验中设定两种情况:一种情况是仅采用有标签训练数据来获得 SFSLs 模型;另一种情况是采用所有的训练数据包括有标签和无标签数据来获得 SFSLs 模型.实验比较了两种情况的标注结果,从而得出无标签数据对标注性能的影响.有标签训练数据大小分为 5 种情况,分别为训练数据总量的 5%,10%,25%,50%和 100%.另外,实验中还将本文方法 SFSLs 与方法 SFSS 进行了比较,分别在两个数据库上进行了实验,实验结果如图 1 所示.

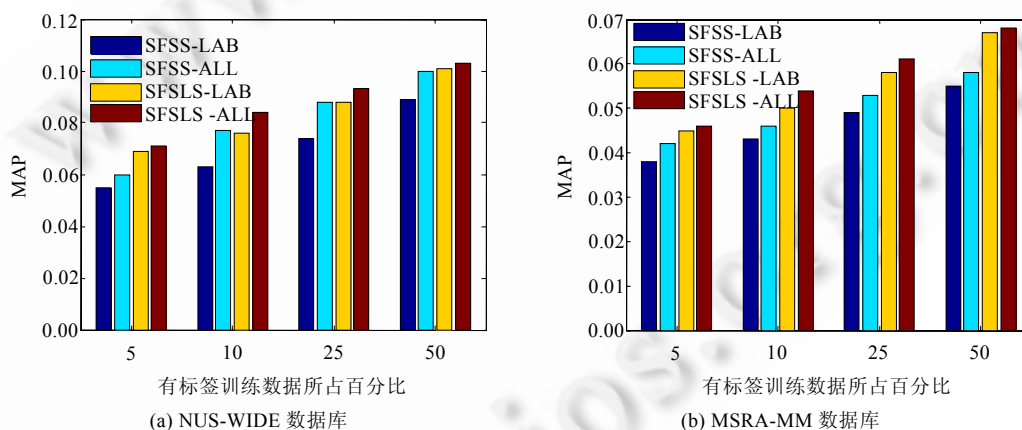


Fig.1 Influence of unlabeled data on annotation results

图 1 无标签数据对标注性能的影响

图 1 给出了无标签数据对标注性能的影响结果,其中,图 1(a)所示为数据库 NUS-WIDE 上的标注结果,图 1(b)所示为数据库 MSRA-MM 上的标注结果.图中相邻的 4 个条形从左到右依次表示仅仅采用有标签数据的 SFSS 方法的标注结果,同时采用有标签数据和无标签数据的 SFSS 方法的标注结果,仅仅采用有标签数据的 SFSLs 方法的标注结果,同时采用有标签数据和无标签数据的 SFSLs 方法的标注结果.

图 1 表明:无论是本文所提算法 SFSLs,还是算法 SFSS,无标签数据的加入能够提升网络图像标注的性能,采用包括有标签和无标签所有训练数据的标注结果要优于仅采用有标签数据的标注结果.同时也显示



出:SFSLs 的性能优于算法 SFSS,无论是仅采用有标签数据的情况,还是采用全部训练数据的情况.

#### 4 结 论

本文提出一个增强稀疏性特征选择模型 SFSLs 进行网络图像标注,该模型基于  $l_{2,1/2}$  矩阵范数进行稀疏特征选择,使选取的特征更为稀疏,更具判别性.算法中引入共享子空间学习,考虑了不同特征之间的相关性.另外,采用基于图拉普拉斯半监督学习方法,同时利用有标签数据和无标签数据,一方面能够节省标注大规模训练数据的时间和人力,同时,易于获取的无标签数据可以提升网络图像标注的性能.SFSLs 算法的目标函数是非凸的,我们设计了一种有效的迭代算法来求取其最优解,并证明了该算法的收敛性.在两个大的网络图像数据库上进行了实验,结果表明:本文算法 SFSLs 优于其他稀疏特征选择算法,更适合大规模网络图像的标注.

**致谢** 在此,我们向对本文的工作给予支持和建议的同行,尤其是北京交通大学信息科学研究所袁保宗教授、郭松、王占等同学表示感谢.

#### References:

- [1] Wu F, Yuan Y, Zhuang YT. Heterogeneous feature selection by group lasso with logistic regression. In: Proc. of the Int'l Conf. on Multimedia. 2010. 983–986. [doi: 10.1145/1873951.1874129].
- [2] Yang Y, Huang Z, Yang Y, Liu JJ, Shen HT, Luo J. Local image tagging via graph regularized joint group sparsity. Pattern Recognition, 2013,46(5):1358–1368. [doi: 10.1016/j.patcog.2012.10.026]
- [3] Yuan Y, Shao J, Wu F, Zhuang YT. Image annotation by the multiple kernel learning with group sparsity effect. Ruan Jian Xue Bao/Journal of Software, 2012,23(9):2500–2509 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4154.htm> [doi: 10.3724/SP.J.1001.2012.04154]
- [4] Ma ZG, Nie FP, Yang Y, Uijlings J, Sebe N, Hauptmann A. Discriminating joint feature analysis for multimedia data understanding. IEEE Trans. on Multimedia, 2012,14(6):1662–1672. [doi: 10.1109/TMM.2012.2199293]
- [5] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 1996,58:267–288.
- [6] Cai D, Zhang C, He X. Unsupervised feature selection for multi-cluster data. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2010. 333–342. [doi: 10.1145/1835804.1835848]
- [7] Xu ZB, Chang XY, Xu FM, Zhang H.  $L_{1/2}$  regularization: A thresholding representation theory and a fast solver. IEEE Trans. on Neural Networks and Learning Systems, 2012,23(7):1013–1027. [doi: 10.1109/TNNLS.2012.2197412]
- [8] Xu ZB, Zhang H, Wang Y, Chang XY, Liang Y.  $L_{1/2}$  regularizer. Science China, 2010,53(6):1159–1169. [doi: 10.1007/s11432-010-0090-0]
- [9] Zhao Q, Meng DY, Xu ZB.  $L_{1/2}$  regularized logistic regression. Pattern Recognition and Artificial Intelligence, 2012,25(5): 721–728 (in Chinese with English abstract).
- [10] Nie F, Huang H, Cai X, Ding C. Efficient and robust feature selection via joint  $L_{2,1}$ -norms minimization. In: Proc. of the Neural Information Processing Systems. 2010. 1813–1821. <http://papers.nips.cc/book/advances-in-neural-information-processing-systems-23-2010>
- [11] Ando RK, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data. The Journal of Machine Learning Research, 2005,6:1817–1853.
- [12] Yang Y, Wu F, Nie FP, Shen HT, Zhuang YT, Hauptmann AG. Web and personal image annotation by mining label correlation with relaxed visual graph embedding. IEEE Trans. on Image Processing. 2012,21(3):1339–1351. [doi: 10.1109/TIP.2011. 2169269]
- [13] Ma ZG, Nie FP, Yang Y, Uijlings JRR, Sebe N. Web image annotation via subspace-sparsity collaborated feature selection. IEEE Trans. on Multimedia, 2012,14(4):1021–1030. [doi: 10.1109/TMM.2012.2187179]
- [14] Tang JH, Hong RC, Yan SC, Chua TS, Qi GJ, Jain R. Image annotation by  $k$ NN-sparse graph-based label propagation over noisily-tagged Web images. ACM Trans. on Intelligent Systems and Technology, 2010,1(1):111–126. [doi: 10.1145/1899412. 1899418]

- [15] Lee WY, Hsieh LC, Wu GL, Hsu W. Graph-Based semi-supervised learning with multi-modality propagation for large-scale image datasets. *Journal of Visual Communication and Image Representation*, 2013,24(3):295–302. [doi: org/10.1016/j.jvcir.2012.12.002]
- [16] Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng Y. Nus-Wide: A real-world Web image database from national university of Singapore. In: *Proc. of the Int'l Conf. on Image and Video Retrieval*. Santorini: ACM Press, 2009. 1–9. [doi: 10.1145/1646396.1646452]
- [17] Li H, Wang M, Hua X. MSRA-MM 2.0: A large-scale Web multimedia dataset. In: *Proc. of the 2009 IEEE Int'l Conf. on Data Mining Workshops*. Miami: IEEE, 2009. 164–169. [doi: 10.1109/ICDMW.2009.46]
- [18] Wang LP, Chen SC.  $l_{2,p}$ -Matrix norm and its application in feature selection. 2013. <http://arxiv.org/abs/1303.3987>
- [19] Zhu X. Semi-Supervised learning literature survey. Technical Report, 1530, Madison: University of Wisconsin, 2007.
- [20] Zhu X, Ghahramani Z, Lafferty J. Semi-Supervised learning using gaussian fields and harmonic functions. In: *Proc. of the 20th Int'l Conf. on Machine Learning*. Washington, 2003. <http://www.machinelearning.org/icml.html>
- [21] Nie FP, Xu D, Hung T, Zhang CS. Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction. *IEEE Trans. on Image Processing*, 2010,19(7):1921–1932. [doi: 10.1109/TIP.2010.2044958]
- [22] Yang Y, Zhuang YT, Wu F, Pan YH. Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval. *IEEE Trans. on Multimedia*, 2008,10(3):437–446. [doi: 10.1109/TMM.2008.917359]
- [23] Zha ZJ, Mei T, Wang JD, Wang ZF, Hua XS. Graph-Based semi-supervised learning with multiple labels. *Journal of Visual Communication and Image Representation*, 2009,20(2):97–103. [doi: 10.1016/j.jvcir.2008.11.009]
- [24] Wang M, Hua XS, Tang JH, Hong R. Beyond distance measurement: Constructing neighborhood similarity for video annotation. *IEEE Trans. on Multimedia*, 2009,11(3):465–476. [doi: 10.1109/TMM.2009.2012919]
- [25] Chen J, Tang L, Liu J, Ye J. A convex formulation for learning shared structures from multiple tasks. In: *Proc. of the 26th Annual Int'l Conf. on Machine Learning*. New York: ACM Press, 2009. 137–144. [doi: 10.1145/1553374.1553392]
- [26] Ji S, Tang L, Yu S, Ye J. A shared-subspace learning framework for multi-label classification. *ACM Trans. on Knowledge Discovery from Data*, 2010,4(2):1–29. [doi: 10.1145/1754428.1754431]
- [27] Golub GH, Loan CF. *Matrix Computations*. 3rd ed., Baltimore: The Johns Hopkins University Press, 1996.
- [28] Cawley G, Talbot N, Girolami M. Sparse multinomial logistic regression via bayesian  $L_1$  regularisation. In: Schölkopf B, Platt J, Hoffman T, eds. *Proc. of the Neural Information Processing Systems*. 2007. 209–216. <http://papers.nips.cc/book/advances-in-neural-information-processing-systems-20-2007>

#### 附中文参考文献:

- [3] 袁莹,邵健,吴飞,庄越挺. 结合组稀疏效应和多核学习的图像标注. *软件学报*,2012,23(9):2500–2509. <http://www.jos.org.cn/1000-9825/4154.htm> [doi: 10.3724/SP.J.1001.2012.04154]
- [9] 赵谦,孟德宇,徐宗本.  $L_{1/2}$  正则化 Logistic 回归. *模式识别与人工智能*,2012,25(5):721–728.

#### 附录

**引理 1.** 设  $G_t$  是第  $t$  次迭代得到的映射矩阵,  $G_{t+1}$  是第  $t+1$  次迭代得到的映射矩阵. 同时, 设  $g_t^i$  和  $g_{t+1}^i$  分别是  $G_t$  和  $G_{t+1}$  的第  $i$  行 ( $i=1,2,\dots,d$ ), 则下面的不等式成立:

$$\sum_{i=1}^d \left( \|g_{t+1}^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_{t+1}^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right) \leq \sum_{i=1}^d \left( \|g_t^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_t^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right).$$

证明: 对于表达式  $\varphi(t)=4t-t^4-3$ , 对于任意的  $t>0$ , 都有  $\varphi(t)\leq 0$ .

令  $\varphi(t)$  中的  $t = \frac{\|g_{t+1}^i\|_2^{1/2}}{\|g_t^i\|_2^{1/2}}$ , 则  $\varphi\left(\frac{\|g_{t+1}^i\|_2^{1/2}}{\|g_t^i\|_2^{1/2}}\right) \leq 0$ , 也即  $4 \frac{\|g_{t+1}^i\|_2^{1/2}}{\|g_t^i\|_2^{1/2}} - \frac{\|g_{t+1}^i\|_2^2}{\|g_t^i\|_2^2} - 3 \leq 0$ .

在上面不等式两侧同时乘以  $\frac{1}{4} \|g_t^i\|_2^{1/2}$ , 则有:

$$\|g_{t+1}^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_{t+1}^i\|_2^2}{\|g_t^i\|_2^{3/2}} \leq \|g_t^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_t^i\|_2^2}{\|g_t^i\|_2^{3/2}}, i=1, \dots, d.$$

上面不等式是针对于每一行  $i$ , 将所有的行相加( $1 \leq i \leq d$ ), 则得到引理 1 的结论:

$$\sum_{i=1}^d \left( \|g_{t+1}^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_{t+1}^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right) \leq \sum_{i=1}^d \left( \|g_t^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_t^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right). \quad \square$$

定理 1. 采用算法 1 中的迭代过程, 使公式(22)中目标函数在每次迭代过程中是单调递减的, 最终收敛于最优的映射矩阵  $G$ .

证明: 根据算法 1, 由公式(22)可以推得:

$$\left. \begin{aligned} G_{t+1} = \arg \min_{F, G, U, P} & Tr(F^T LF) + Tr((F - Y)^T Q(F - Y)) + \\ & \alpha [Tr((X^T G - F)^T (X^T G - F)) + 4\lambda Tr(G^T DG) + \mu Tr((G - UP)^T (G - UP))] \\ \text{s.t. } & U^T U = I \end{aligned} \right\}$$

因此:

$$\begin{aligned} & Tr(F_{t+1}^T L F_{t+1}) + Tr((F_{t+1} - Y)^T Q(F_{t+1} - Y)) + \\ & \alpha [Tr((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda Tr(G_{t+1}^T D_t G_{t+1}) + \mu Tr((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1}))] \leq \\ & Tr(F_t^T L F_t) + Tr((F_t - Y)^T Q(F_t - Y)) + \\ & \alpha [Tr((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda Tr(G_t^T D_t G_t) + \mu Tr((G_t - U_t P_t)^T (G_t - U_t P_t))]. \\ \therefore Tr(G_t^T D_t G_t) &= \sum_{i=1}^d \frac{\|g_t^i\|_2^2}{4 \|g_t^i\|_2^{3/2}}, Tr(G_{t+1}^T D_t G_{t+1}) = \sum_{i=1}^d \frac{\|g_{t+1}^i\|_2^2}{4 \|g_t^i\|_2^{3/2}}, \\ \therefore Tr(F_{t+1}^T L F_{t+1}) + Tr((F_{t+1} - Y)^T Q(F_{t+1} - Y)) + \\ & \alpha \left[ Tr((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda \sum_{i=1}^d \frac{\|g_{t+1}^i\|_2^2}{4 \|g_t^i\|_2^{3/2}} + \mu Tr((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1})) \right] \leq \\ & Tr(F_t^T L F_t) + Tr((F_t - Y)^T Q(F_t - Y)) + \\ & \alpha \left[ Tr((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda \sum_{i=1}^d \frac{\|g_t^i\|_2^2}{4 \|g_t^i\|_2^{3/2}} + \mu Tr((G_t - U_t P_t)^T (G_t - U_t P_t)) \right] \\ \Rightarrow Tr(F_{t+1}^T L F_{t+1}) + Tr((F_{t+1} - Y)^T Q(F_{t+1} - Y)) + & \alpha \left[ Tr((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda \sum_{i=1}^d \|g_{t+1}^i\|_2^{1/2} - \right. \\ & \left. 4\lambda \sum_{i=1}^d \|g_{t+1}^i\|_2^{1/2} + 4\lambda \sum_{i=1}^d \frac{\|g_{t+1}^i\|_2^2}{4 \|g_t^i\|_2^{3/2}} + \mu Tr((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1})) \right] \leq \\ & Tr(F_t^T L F_t) + Tr((F_t - Y)^T Q(F_t - Y)) + \alpha \left[ Tr((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda \sum_{i=1}^d \|g_t^i\|_2^{1/2} - 4\lambda \sum_{i=1}^d \|g_t^i\|_2^{1/2} + \right. \\ & \left. 4\lambda \sum_{i=1}^d \frac{\|g_t^i\|_2^2}{4 \|g_t^i\|_2^{3/2}} + \mu Tr((G_t - U_t P_t)^T (G_t - U_t P_t)) \right] \\ \Rightarrow Tr(F_{t+1}^T L F_{t+1}) + Tr((F_{t+1} - Y)^T Q(F_{t+1} - Y)) + & \alpha \left[ Tr((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda \sum_{i=1}^d \|g_{t+1}^i\|_2^{1/2} - \right. \\ & \left. 4\lambda \sum_{i=1}^d \left( \|g_{t+1}^i\|_2^{1/2} - \frac{\|g_{t+1}^i\|_2^{22}}{4 \|g_t^i\|_2^{3/2}} \right) + \mu Tr((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1})) \right] \leq \\ & Tr(F_t^T L F_t) + Tr((F_t - Y)^T Q(F_t - Y)) + \alpha \left[ Tr((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda \sum_{i=1}^d \|g_t^i\|_2^{1/2} - \right. \\ & \left. 4\lambda \sum_{i=1}^d \left( \|g_t^i\|_2^{1/2} - \frac{\|g_t^i\|_2^2}{4 \|g_t^i\|_2^{3/2}} \right) + \mu Tr((G_t - U_t P_t)^T (G_t - U_t P_t)) \right]. \end{aligned}$$

引理 1 中已经证明:

$$\begin{aligned}
& \sum_{i=1}^d \left( \|g_{t+1}^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_{t+1}^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right) \leq \sum_{i=1}^d \left( \|g_t^i\|_2^{1/2} - \frac{1}{4} \frac{\|g_t^i\|_2^2}{\|g_t^i\|_2^{3/2}} \right) \\
\Rightarrow & \text{Tr}(F_{t+1}^T L F_{t+1}) + \text{Tr}((F_{t+1} - Y)^T Q (F_{t+1} - Y)) + \\
& \alpha \left[ \text{Tr}((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda \sum_{i=1}^d \|g_{t+1}^i\|_2^{1/2} + \mu \text{Tr}((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1})) \right] \leq \\
& \text{Tr}(F_t^T L F_t) + \text{Tr}((F_t - Y)^T Q (F_t - Y)) + \\
& \alpha \left[ \text{Tr}((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda \sum_{i=1}^d \|g_t^i\|_2^{1/2} + \mu \text{Tr}((G_t - U_t P_t)^T (G_t - U_t P_t)) \right], \\
\Rightarrow & \text{Tr}(F_{t+1}^T L F_{t+1}) + \text{Tr}((F_{t+1} - Y)^T Q (F_{t+1} - Y)) + \\
& \alpha \left[ \text{Tr}((X^T G_{t+1} - F_{t+1})^T (X^T G_{t+1} - F_{t+1})) + 4\lambda \|G_{t+1}\|_{2,1/2}^{1/2} + \mu \text{Tr}((G_{t+1} - U_{t+1} P_{t+1})^T (G_{t+1} - U_{t+1} P_{t+1})) \right] \leq \\
& \text{Tr}(F_t^T L F_t) + \text{Tr}((F_t - Y)^T Q (F_t - Y)) + \\
& \alpha \left[ \text{Tr}((X^T G_t - F_t)^T (X^T G_t - F_t)) + 4\lambda \|G_t\|_{2,1/2}^{1/2} + \mu \text{Tr}((G_t - U_t P_t)^T (G_t - U_t P_t)) \right].
\end{aligned}$$

上式表明:根据算法 1,目标函数(22)是单调递减的,最终收敛于最优映射矩阵  $G$ . □



史彩娟(1977—),女,河北丰润人,博士生,讲师,主要研究领域为多媒体分析,计算机视觉,机器学习.



阮秋琦(1944—),男,教授,博士生导师,主要研究领域为图像处理,模式识别.