基于对偶超图正则化的概念分解算法及其在数据表示中的应用

叶 军[1] 金 忠[2]

(南京邮电大学理学院 南京 210046)[1] (南京理工大学计算机科学与工程学院 南京 210094)[2]

摘 要 针对概念分解算法没有同时考虑数据空间和特征属性空间中的高阶几何结构信息的问题,提出了一种基于对偶超图正则化的概念分解算法。该算法通过分别在数据空间和特征属性空间中构建无向加权的拉普拉斯超图正则项,分别反映了数据流形和特征流形的多元几何结构信息,弥补了传统图模型只能表达数据间成对关系的缺陷。采用交替迭代的方法求解算法的目标函数并证明了算法的收敛性。在3个真实数据库(TDT2、PIE、COIL20)上的实验表明,该方法在数据的聚类表示的效果方面优于其他方法。

关键词 概念分解,超图学习,对偶回归,流形学习,聚类

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.07.056

Hypergraph Dual Regularization Concept Factorization Algorithm and Its Application in Data Representation

YE Jun[1] JIN Zhong[2]

(School of Natural Sciences, Nanjing University of Posts & Telecommunications, Nanjing 210046, China)[1]

(School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094, China)[2]

Abstract The concept factorization(CF) algorithm can not take the geometric structures of both the data manifold and the feature manifold into account simultaneously. And CF algorithm can not consider the high-order relationship among samples. In this paper, a novel algorithm called hypergraph dual regularization concept factorization(DHCF) algorithm was proposed, which encodes the high-order geometric structure information of data and feature spaces by constructing two undirected weighted hypergraph Laplacian regularize term, respectively. By this way, the proposed method can overcome the deficiency that traditional graph model expresses pair-wise relationship only. Moreover, we developed the iterative updating optimization schemes for DHCF, and provided the convergence proof of our optimization scheme. Experimental results on TDT2 document datasets, PIE and COIL20 image datasets demonstrate the effectiveness of our method.

Keywords CF, Hypergraph learning, Dual regularized, Manifold learning, Clustering

1 引言

随着计算机及网络技术的发展,人们可获取越来越多的高维数据,如文本、图像数据等。如何从海量的高维数据中挖掘出隐藏信息和有效数据,已成为现今机器学习、数据挖掘、社会网络分析等领域的研究热点[1-3]。合适的数据表示方式能够挖掘出数据中的潜在结构,有利于数据的进一步处理。目前,矩阵分解方法作为一种有效的数据处理方式引起了许多研究者的关注。常用的矩阵分解算法包括奇异值分解(Singular Value Decomposition, SVD)、非负矩阵分解(Nonnegative Matrix Factorization, NMF)[4]和概念分解(Concept Factorization, CF)[5]等。

最近的研究表明,人们观测到的数据一般都分布在高维空间中的低维子流形上,并有大量的流形学习技术被提出,用于发现数据潜在的几何结构[6],如ISOMAP[7],LLE[8],Laplace[9]特征映射等。这些流形学习方法的基本思想都是保持局部不变性,即近邻点应该有相似的数据表示。在NMF和CF的框架下,已有许多文献对其进行了拓展。结合标签信息和稀疏性约束,胡学考等[10]提出了基于稀疏约束的半监督非负矩阵分解算法(Nonnegative Matrix Factorization with Sparseness Constraints, NMFSC)。方蔚涛等[11]将二维主成分分析(2DPCA)的思想和NMF的思想结合起来,提出了二维投影非负矩阵分解(2DPNMF)算法,该方法直接针对二维图像,不需要先将图像转化为一个向量,很好地保留了图像行列之间的局部相关关系。通过将图学习思想融入NMF和CF方法中[12-15],文献[12]提出了图正则的非负矩阵分解(Graph regularized NMF, GNMF)算法,文献[15]提出了局部连续概念分解算法(Locally Consistent CF, LCCF),这两种方法在文本聚类及人脸识别等应用中均取得了不错的效果,进一步表明了数据的空间结构信息可以有效地提高学习的质

到稿日期:2016-06-11 返修日期:2016-08-01 本文受国家自然科学基金项目(61373063),江苏省自然科学基金项目(BK20150867),南京邮电大学国家自然科学基金孵化项目(NY215125)资助。

叶 军(1981-),男,博士,副教授,主要研究方向为模式识别、机器学习、图像处理,E-mail:yj8422092@163.com;金 忠(1961-),男,博士,教授,博士生导师,主要研究方向为模式识别、人脸识别、机器学习、计算机视觉。

量。然而这些工作只考虑了数据空间的分布结构，没有利用特征属性空间的结构信息。最近，Shang F 等[6]和 Ye J 等[16]分别在非负矩阵分解和概念分解框架下提出了同时考虑数据流形和特征流形的几何结构的双图正则化非负矩阵分解(Graph Dual regularization Nonnegative Matrix Factorization，DNMF)算法和双图正则化概念分解(Graph Dual regularization Concept Factorization，GCF)算法，均取得了不错的效果。这也进一步说明了不但数据空间的几何结构信息可以有效地提高学习的质量，同时特征属性空间的几何结构信息也能对学习的质量起到辅助作用。但这些方法均使用简单图学习的方式来反映数据对象间的局部几何关系。事实上，实际中的数据分布是非常复杂的，简单图学习不能挖掘出数据间更高阶的几何关系，从而不能更好地利用几何信息来对数据进行处理表示。超图学习[17-18]扩展了传统简单图模型中两个顶点组建边的构图方式，以具有某种相似属性的数据子集构建超边，可以有效地刻画数据间的高阶关系，文献[19-21]表明，将超图学习用于矩阵分解中可以大大提升分类性能。

因此，为了进一步挖掘数据间更高阶的几何信息，在双图正则化概念分解(GCF)算法的基础上，提出了一种基于对偶超图正则化的概念分解算法(Hypergraph dual regularization Concept Factorization，DHCF)，该算法分别在数据空间和特征空间构建超图来反映它们各自的分布流形几何结构信息，以此来获得样本间高阶的几何结构信息。建立基于对偶超图正则化的概念分解模型，推导出该算法的交替迭代更新规则，给出该算法的收敛性证明，实验结果表明了算法的有效性和准确性。

2 相关工作

2.1 概念分解(CF)算法

给定一个非负矩阵 $\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N]\in\boldsymbol{R}^{M\times N}$，$\boldsymbol{X}$ 的每一列代表一个样本。CF 算法的目标是寻求两个非负矩阵 $\boldsymbol{W}=[w_{jk}]\in\boldsymbol{R}^{N\times K}$，$\boldsymbol{V}=[v_{jk}]\in\boldsymbol{R}^{N\times K}$，其中 $K\ll\min\{M,N\}$，使其满足 $\boldsymbol{X}\approx\boldsymbol{X}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}$。CF 的目标函数可表示为：

$$\min_{\boldsymbol{W},\boldsymbol{V}}:\boldsymbol{J}_{\boldsymbol{CF}}=\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}\|_F^2\quad \text{s.t.}\ \boldsymbol{W},\boldsymbol{V}\geqslant 0 \tag{1}$$

针对式(1)，记 $\boldsymbol{K}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$，Xu 等[5]给出了相关的乘积更新迭代算法来求解上述问题。

$$\begin{aligned} w_{jk}^{t+1}&=w_{jk}^{t}\frac{(\boldsymbol{K}\boldsymbol{V})_{jk}}{(\boldsymbol{K}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{jk}}\\ v_{jk}^{t+1}&=v_{jk}^{t}\frac{(\boldsymbol{K}\boldsymbol{W})_{jk}}{(\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{W})_{jk}}\end{aligned} \tag{2}$$

2.2 双图正则化概念分解(GCF)算法

Ye J 等[16]在概念分解算法的基础上同时考虑了数据和特征属性的几何结构信息，提出了双图正则化的概念分解算法。其目标函数可表示为：

$$\min_{\boldsymbol{W},\boldsymbol{V}}:\boldsymbol{J}_{\boldsymbol{GCF}}=\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}\|^2+\alpha\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{V}}\boldsymbol{V})+\beta\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{W}}\boldsymbol{W}) \tag{3}$$

$$\text{s.t.}\ \boldsymbol{W},\boldsymbol{V}\geqslant 0$$

其中，$\boldsymbol{L}_{\boldsymbol{W}}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{U}}\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{D}^{\boldsymbol{U}}-\boldsymbol{S}^{\boldsymbol{U}})\boldsymbol{X}=\boldsymbol{D}^{\boldsymbol{W}}-\boldsymbol{S}^{\boldsymbol{W}}$。

针对式(3)，Ye J 等人[16]给出了相关的乘积更新迭代算法来求解上述问题。

$$\begin{aligned} w_{jk}^{t+1}&\leftarrow w_{jk}^{t}\frac{(\boldsymbol{K}\boldsymbol{V}+\beta\boldsymbol{S}^{\boldsymbol{W}}\boldsymbol{W})_{jk}}{(\boldsymbol{K}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}+\beta\boldsymbol{D}^{\boldsymbol{W}}\boldsymbol{W})_{jk}}\\ v_{jk}^{t+1}&\leftarrow v_{jk}^{t}\frac{(\boldsymbol{K}\boldsymbol{W}+\alpha\boldsymbol{S}^{\boldsymbol{V}}\boldsymbol{V})_{jk}}{(\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{W}+\alpha\boldsymbol{D}^{\boldsymbol{V}}\boldsymbol{V})_{jk}}\end{aligned} \tag{4}$$

3 基于对偶超图正则化的概念分解(DHCF)算法

3.1 构建对偶超图正则项

假设超图 $G=(V_1,E,\boldsymbol{W})$，V_1 是节点集合，E 为超边集合，$\boldsymbol{W}$ 为由超边的权重构成的对角矩阵，其中 $\boldsymbol{W}_{ij}=\sum_{e\in E}\sum_{(i,j)\in e}\frac{w(e)}{\delta(e)}$ 为任意两点的权重。文献[17]给出了超边权重的计算方法，定义 $\boldsymbol{D}_v$ 和 $\boldsymbol{D}_e$ 为对角矩阵，分别表示顶点的度和超边的度。根据文献[17]，超图的拉普拉斯矩阵可表示为 $\boldsymbol{L}^{hyper}=\boldsymbol{D}_v-\boldsymbol{S}$，其中 $\boldsymbol{S}=\boldsymbol{H}\boldsymbol{W}\boldsymbol{D}_e^{-1}\boldsymbol{H}^{\mathrm{T}}$。数据映射到低维空间后，构建超图正则项。最近的研究表明：观测到的数据分布在一个低维子流形上，称为数据流形；数据的特征也分布在一个低维子流形上，称为特征流形[6]。我们分别用两个超图来刻画数据流形和特征流形的几何结构，即数据超图和特征超图。

3.1.1 数据超图

设图的顶点集合为数据集 $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N\}$，根据文献[17-18]计算超边权重，则数据超图的拉普拉斯矩阵可表示为 $\boldsymbol{L}_{\boldsymbol{V}}^{hyper}=\boldsymbol{D}_v^{\boldsymbol{V}}-\boldsymbol{S}^{\boldsymbol{V}}$。令 $\boldsymbol{V}=[\boldsymbol{v}_1^{\mathrm{T}},\cdots,\boldsymbol{v}_N^{\mathrm{T}}]^{\mathrm{T}}\in\boldsymbol{R}^{N\times K}$ 为待求的低维数据表示，则该数据表示的平滑度为：

$$\begin{aligned}\mathcal{O}_1&=\frac{1}{2}\sum_{e\in E}\sum_{(i,j)\in e}\frac{w(e)}{\delta(e)}\|v_i-v_j\|^2\\&=\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{D}_v^{\boldsymbol{V}}\boldsymbol{V})-\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{S}^{\boldsymbol{V}}\boldsymbol{V})\\&=\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{V}}^{hyper}\boldsymbol{V})\end{aligned} \tag{5}$$

3.1.2 特征超图

类似地，将图的顶点集合定义为特征集 $\{\boldsymbol{x}_1^{\mathrm{T}},\cdots,\boldsymbol{x}_M^{\mathrm{T}}\}$，重新构造超图，可以得到特征图的拉普拉斯矩阵：$\boldsymbol{L}_{\boldsymbol{W}}^{hyper}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{U}}^{hyper}\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{D}_v^{\boldsymbol{U}}-\boldsymbol{S}^{\boldsymbol{U}})\boldsymbol{X}=\boldsymbol{D}_v^{\boldsymbol{W}}-\boldsymbol{S}^{\boldsymbol{W}}$，令 $\boldsymbol{W}=[\boldsymbol{w}_1^{\mathrm{T}},\cdots,\boldsymbol{w}_N^{\mathrm{T}}]^{\mathrm{T}}\in\boldsymbol{R}^{N\times K}$ 为待求的低维数据表示，则相应的平滑度为：

$$\begin{aligned}\mathcal{O}_2&=\frac{1}{2}\sum_{e\in E}\sum_{(i,j)\in e}\frac{w(e)}{\delta(e)}\|w_i-w_j\|^2\\&=\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{D}_v^{\boldsymbol{W}}\boldsymbol{W})-\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}^{\boldsymbol{W}}\boldsymbol{W})\\&=\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{W}}^{hyper}\boldsymbol{W})\end{aligned} \tag{6}$$

3.2 构建 DHCF 算法的目标函数

为了同时考虑样本的数据流形和特征流形的几何结构信息，在 CF 算法的目标函数中添加基于超图的数据图和特征图正则项，得到 DHCF 算法的目标函数为：

$$\begin{aligned}&\min_{\boldsymbol{W},\boldsymbol{V}}:\boldsymbol{J}_{\boldsymbol{DHCF}}(\boldsymbol{W},\boldsymbol{V})=\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{W}\boldsymbol{V}^{\mathrm{T}}\|^2+\alpha\mathcal{O}_1+\beta\mathcal{O}_2\\&\text{s.t.}\ \boldsymbol{W},\boldsymbol{V}\geqslant 0\end{aligned} \tag{7}$$

其中，$\alpha\geqslant 0$，$\beta\geqslant 0$ 为正则化参数。

3.3 DHCF 目标函数的求解

DHCF 算法中的目标函数 $\boldsymbol{J}_{\boldsymbol{DHCF}}(\boldsymbol{W},\boldsymbol{V})$ 是关于两个变量 $\boldsymbol{W}$ 和 $\boldsymbol{V}$ 的非凸函数，因此求其全局最优解是不现实的。利用交替迭代法可得到问题的局部最优解。根据矩阵的两个性质：$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{B})=\mathrm{Tr}(\boldsymbol{B}\boldsymbol{A})$ 和 $\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}})=\mathrm{Tr}(\boldsymbol{A})$，记 $\boldsymbol{K}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$，目标函数 $\boldsymbol{J}_{\boldsymbol{DHCF}}(\boldsymbol{W},\boldsymbol{V})$ 可重写为：

$$\begin{aligned}\boldsymbol{J}_{DHCF}&=\mathrm{Tr}[(\boldsymbol{X}-\boldsymbol{XWV}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{XWV}^{\mathrm{T}})]+\alpha\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{V}^{hyper}\boldsymbol{V})+\beta\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{W}^{hyper}\boldsymbol{W})\\&=\mathrm{Tr}(\boldsymbol{K})-2\mathrm{Tr}(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{K})+\mathrm{Tr}(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KWV}^{\mathrm{T}})+\alpha\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{V}^{hyper}\boldsymbol{V})+\beta\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{W}^{hyper}\boldsymbol{W})\end{aligned}\tag{8}$$

令 $\boldsymbol{\Psi}=[\psi_{jk}]$，$\boldsymbol{\Phi}=[\varphi_{jk}]$ 为约束 $\boldsymbol{W}\geqslant 0$ 和 $\boldsymbol{V}\geqslant 0$ 对应的拉格朗日乘子，则式(7)的拉格朗日函数 $\boldsymbol{L}$ 为：

$$\boldsymbol{L}=\mathrm{Tr}(\boldsymbol{K})-2\mathrm{Tr}(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{K})+\mathrm{Tr}(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KWV}^{\mathrm{T}})+\alpha\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{V}^{hyper}\boldsymbol{V})+\beta\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{W}^{hyper}\boldsymbol{W})+\mathrm{Tr}(\boldsymbol{\Psi W}^{T})+\mathrm{Tr}(\boldsymbol{\Phi V})\tag{9}$$

对函数 $\boldsymbol{L}$ 分别关于 $\boldsymbol{W}$ 和 $\boldsymbol{V}$ 求偏导，由 KKT 最优性条件可以得到 DHCF 算法的更新迭代公式为：

$$w_{jk}^{t+1}\leftarrow w_{jk}^{t}\frac{(\boldsymbol{KV}+\beta\boldsymbol{S}^{W}\boldsymbol{W})_{jk}}{(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V}+\beta\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{jk}}\tag{10}$$

$$v_{jk}^{t+1}\leftarrow v_{jk}^{t}\frac{(\boldsymbol{KW}+\alpha\boldsymbol{S}^{V}\boldsymbol{V})_{jk}}{(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KW}+\alpha\boldsymbol{D}_{v}^{V}\boldsymbol{V})_{jk}}\tag{11}$$

3.4　DHCF 算法的收敛性证明

定义 1　当满足条件：$G(w,w')\geqslant F(w)$ 和 $G(w,w)=F(w)$ 时，$G(w,w')$ 为 $F(w)$ 的一个辅助函数。

引理 1　若 G 为 F 的辅助函数，则函数 F 在如下的更新公式下为单调下降的。

$$w^{(K+1)}=\arg\min_{w}G(w,w^{(K)})\tag{12}$$

证明：

$$F(w^{(K+1)})\leqslant G(w^{(K+1)},w^{(K)})\leqslant G(w^{(K)},w^{(K)})=F(w^{(K)})$$

令 w_{ab} 为矩阵 $\boldsymbol{W}$ 的元素，$F_{w_{ab}}$ 为目标函数 $\boldsymbol{J}_{DHCF}$ 中仅与 w_{ab} 元素有关的项，即：

$$F_{w_{ab}}=\|\boldsymbol{X}-\boldsymbol{XWV}^{\mathrm{T}}\|^{2}+\beta\mathrm{Tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{L}_{W}^{hyper}\boldsymbol{W})\tag{13}$$

考虑到算法是基于元素运算的，首先证明 $F_{w_{ab}}$ 在式(10)下为单调下降的。事实上，由于：

$$F'_{w_{ab}}=(-2\boldsymbol{KV}+2\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V}+2\beta\boldsymbol{L}_{W}^{hyper}\boldsymbol{W})_{ab}$$

$$F''_{w_{ab}}=2(\boldsymbol{K})_{aa}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}+2\beta(\boldsymbol{L}_{W}^{hyper})_{aa}$$

引理 2　函数 $G(w,w_{ab}^{(K)})$ 为 $F_{w_{ab}}$ 的辅助函数。

$$G(w,w_{ab}^{(K)})=F_{w_{ab}}(w_{ab}^{(K)})+F'_{w_{ab}}(w_{ab}^{(K)})(w-w_{ab}^{(K)})+\frac{(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V})_{ab}+\beta(\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{ab}}{w_{ab}^{(K)}}(w-w_{ab}^{(K)})^{2}\tag{14}$$

证明：由定义 1，显然 $G(w,w)=F_{w_{ab}}(w)$。令 $F_{w_{ab}}(w)$ 的 Taylor 展开序列为：

$$F_{w_{ab}}(w)=F_{w_{ab}}(w_{ab}^{(K)})+F'_{w_{ab}}(w_{ab}^{(K)})(w-w_{ab}^{(K)})+[(\boldsymbol{K})_{aa}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}+\beta(\boldsymbol{L}_{W}^{hyper})_{ab}](w-w_{ab}^{(K)})^{2}$$

由式(14)可知，证明 $G(w,w_{ab}^{(K)})\geqslant F_{w_{ab}}(w)$ 等价于证明

$$\frac{(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V})_{ab}+\beta(\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{ab}}{w_{ab}^{(K)}}\geqslant(\boldsymbol{K})_{aa}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}+\beta(\boldsymbol{L}_{W}^{hyper})_{aa}\tag{15}$$

事实上：

$$(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V})_{ab}=\sum_{l=1}^{k}(\boldsymbol{KW})_{al}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{lb}\geqslant(\boldsymbol{KW})_{ab}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}\geqslant\sum_{l=1}^{k}(\boldsymbol{K})_{al}w_{lb}^{(K)}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}\geqslant w_{ab}^{(K)}(\boldsymbol{K})_{aa}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{bb}$$

且

$$\beta(\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{ab}=\beta\sum_{j=1}^{M}(\boldsymbol{D}_{v}^{W})_{aj}w_{jb}^{(K)}\geqslant\beta(\boldsymbol{D}_{v}^{W})_{aa}w_{ab}^{(K)}\geqslant\beta(\boldsymbol{D}_{v}^{W}-\boldsymbol{S}^{W})_{aa}w_{ab}^{(K)}=\beta(\boldsymbol{L}_{W}^{hyper})_{aa}w_{ab}^{(K)}$$

因此，不等式(15)成立，即 $G(w,w_{ab}^{(K)})\geqslant F_{w_{ab}}(w)$。

引理 3　函数 $G(v,v_{ab}^{(t)})$ 为 $F_{v_{ab}}$ 的辅助函数。

$$G(v,v_{ab}^{(t)})=F_{v_{ab}}(v_{ab}^{(t)})+F'_{v_{ab}}(v_{ab}^{(t)})(v-v_{ab}^{(t)})+\frac{(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KW})_{ab}+\alpha(\boldsymbol{D}_{v}^{V}\boldsymbol{V})_{ab}}{v_{ab}^{(t)}}(v-v_{ab}^{(t)})^{2}\tag{16}$$

其中，$F_{v_{ab}}=\|\boldsymbol{X}-\boldsymbol{XWV}^{\mathrm{T}}\|^{2}+\alpha\mathrm{Tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}_{V}^{hyper}\boldsymbol{V})$。

引理 3 的证明过程同引理 2 的证明，限于篇幅，此处具体证明参见引理 2。

定理 1　对于给定的数据 $\boldsymbol{X}$ 及任意的 $\boldsymbol{W}\geqslant 0$，$\boldsymbol{V}\geqslant 0$，提出的交替迭代更新规则式(10)、式(11)可使得目标函数 $\boldsymbol{J}_{DHCF}(\boldsymbol{W},\boldsymbol{V})$ 单调下降。

证明：由引理 2，将式(14)、式(16)分别代入式(12)得：

$$w_{ab}^{(K+1)}=w_{ab}^{(K)}-w_{ab}^{(K)}\frac{F'_{ab}(w_{ab}^{(K)})}{2(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V})_{ab}+2\beta(\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{ab}}=w_{ab}^{(K)}\frac{(\boldsymbol{KV}+\beta\boldsymbol{S}^{W}\boldsymbol{W})_{ab}}{(\boldsymbol{KWV}^{\mathrm{T}}\boldsymbol{V}+\beta\boldsymbol{D}_{v}^{W}\boldsymbol{W})_{ab}}\tag{17}$$

$$v_{ab}^{(t+1)}=v_{ab}^{(t)}-v_{ab}^{(t)}\frac{F'_{ab}(v_{ab}^{(t)})}{2(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KW})_{ab}+2\alpha(\boldsymbol{D}_{v}^{V}\boldsymbol{V})_{ab}}=v_{ab}^{(t)}\frac{(\boldsymbol{KW}+\alpha\boldsymbol{S}^{V}\boldsymbol{V})_{ab}}{(\boldsymbol{VW}^{\mathrm{T}}\boldsymbol{KW}+\alpha\boldsymbol{D}_{v}^{V}\boldsymbol{V})_{ab}}\tag{18}$$

由于式(14)、式(16)分别为 $F_{w_{ab}}$ 及 $F_{v_{ab}}$ 的辅助函数，因此 $F_{w_{ab}}$ 及 $F_{v_{ab}}$ 分别在迭代更新式(10)、式(11)下是单调下降的。

3.5　DHCF 算法的具体步骤

输入：数据集矩阵 **X**

输出：分解后的矩阵 **W** 和 **V**

1. 初始化参数，设定近邻点参数 p、分解维度 K、正则化参数 α 和 β，以及最大迭代次数(IterMax)；
2. 计算数据超图和特征超图的邻接矩阵 $\mathbf{S}^{V}$ 和 $\mathbf{S}^{W}$；
3. 计算对角矩阵 $\mathbf{D}_{v}^{V}$ 和 $\mathbf{D}_{v}^{W}$；
4. 随机生成非负矩阵 **W** 和 **V**；
   For t=1：IterMax(t 为迭代次数)
5. 固定 $\mathbf{V}^{(t)}$，根据式(10)更新 **W** 得 $\mathbf{W}^{(t+1)}$；
6. 固定 $\mathbf{W}^{(t+1)}$，根据式(11)更新 **V** 得 $\mathbf{V}^{(t+1)}$；
7. 重复步骤 2 和步骤 3，直至目标函数式(7)收敛；
   End For

3.6　DHCF 算法的复杂度分析

为了比较本文所提算法和其他算法的复杂度，对 CF，LCCF，GCF 和 DHCF 算法的计算复杂度进行了比对。经过 t 次迭代更新后，NMF 算法的计算复杂度为 $O(tMNK)$。CF，LCCF 和 GCF 算法的计算复杂度分别为 $O(tN^{2}K+N^{2}M)$，$O(N^{2}M+tN^{2}K)$ 和 $O(N^{2}M+NM^{2}+tN^{2}K)$。对于 DHCF 算法而言，由于创建数据超图和特征超图所需的计算复杂度为 $O(N^{2}M+NM^{2})$，且数据超图和特征超图中的加权矩阵 $\boldsymbol{S}^{W}$ 和 $\boldsymbol{S}^{V}$ 都是稀疏矩阵，因此迭代求解的计算复杂度为 $O(tN^{2}K)$，故本文所提算法的总计算复杂度为 $O(N^{2}M+NM^{2}+tN^{2}K)$。

4　数值实验

4.1　聚类实验

为了验证本文所提算法的有效性，分别在文本数据库 TDT2 以及图像数据库 PIE 和 COIL20 这 3 个数据集上将所提DHCF算法与 NMF，GNMF，CF，LCCF，GCF 算法进行聚类比较实验。聚类实验中常用准确率(ACCuracy，ACC)和归

一化互信息(Normalized Mutual Information,NMI)[5]作为聚类算法的评价标准。

(1)TDT2 文本数据库:包含了 56 类的 10021 个文档。

(2)PIE 人脸数据库:选择了 68 人的 11554 幅图像,除第 38 个人是 164 幅图像外,其余人都是 170 幅图像。本实验中图像是大小为 32 * 32 的灰度图像。

(3)COIL20 物体图像数据库:包含了 20 个物体的 1440 幅图像,图像是大小为 32 * 32 的灰度图像。

实验中,对于不同的聚类数 $k(k=2,\cdots,10)$,DHCF 算法与 5 种相关的比较算法在 3 个数据库上进行 20 次实验的平均聚类结果如表 1—表 3 所列。从表 1—表 3 可得到如下的结论:

(1)在 TDT2 数据库中,本文所提算法比传统 CF 算法的平均 ACC 和 NMI 分别提高了 11.38%和 12.21%,比 GCF 算法分别提高了 2.09%和 3.19%。在 PIE 数据库中,DHCF 算法比 CF 算法的平均 ACC 和 NMI 分别提高了 13.78%和 12.08%,比 GCF 算法分别提高了 3.08%和 2.35%。在 COIL20 数据库中,DHCF 算法比 CF 算法的平均 ACC 和 NMI 分别提高了 11.95%和 14.02%,比 GCF 算法分别提高了4.38%和4.86%。

(2)GNMF 和 LCCF 算法由于利用了数据分布的几何结构信息,所取得的聚类效果比 NMF 和 CF 本身所得效果更好。这表明数据分布的几何结构在进行聚类时是有效的,特别对于图像数据集,它们的空间分布具有潜在的流形结构。另外,在 3 个数据库中 GCF 算法所得效果优于 LCCF 算法,这是因为 GCF 算法不但考虑了数据流形的几何结构信息,而且也利用了特征流形的结构信息。

(3)最后,本文提出的 DHCF 算法比 GCF 算法的聚类效果更好,其原因是 DHCF 算法在构造反映数据流形和特征流形最本质的几何结构信息时,利用超图进行构造,能够挖掘出样本间更高阶的信息,尤其是在图像数据集中反映得更加的明显。

表 1 TDT2 数据库上的聚类实验结果

k	Accuracy/%						Normalized Mutual Information/%					
	NMF	GNMF	CF	LCCF	GCF	DHCF	NMF	GNMF	CF	LCCF	GCF	DHCF
2	85.69	93.32	86.13	94.23	96.21	97.43	65.72	83.79	67.28	84.18	85.94	86.67
3	79.51	87.29	79.67	88.53	91.04	92.38	63.68	79.82	68.91	78.23	81.33	83.72
4	76.58	85.76	78.29	86.65	89.45	90.19	68.70	76.13	69.78	77.58	80.64	82.38
5	69.34	83.11	73.63	83.41	87.66	88.25	62.35	70.89	64.65	72.04	76.56	80.14
6	71.43	80.97	74.82	81.26	84.61	86.46	67.51	71.99	67.86	74.86	77.41	81.78
7	69.12	78.34	70.53	79.12	80.39	82.25	65.87	70.32	66.54	73.32	75.17	79.69
8	66.80	73.32	67.04	74.28	75.20	79.87	67.89	68.31	68.31	69.73	72.58	77.12
9	67.23	74.47	67.35	75.18	73.28	77.32	67.73	69.26	69.43	70.42	72.20	76.06
10	66.48	69.01	68.65	69.24	71.86	74.41	66.96	68.31	69.81	69.59	71.85	74.83
Avg.	72.46	80.62	74.01	81.32	83.30	85.39	66.27	73.20	68.06	74.44	77.08	80.27

表 2 PIE 数据库上的聚类实验结果

k	Accuracy/%						Normalized Mutual Information/%					
	NMF	GNMF	CF	LCCF	GCF	DHCF	NMF	GNMF	CF	LCCF	GCF	DHCF
2	56.24	61.31	57.23	62.89	67.14	69.25	47.67	57.83	48.62	60.43	64.59	66.14
3	53.37	58.56	58.14	58.42	69.25	72.59	44.83	55.67	49.14	57.85	61.78	64.55
4	54.53	60.04	58.36	60.12	71.71	74.30	44.46	52.76	47.23	53.54	58.38	61.02
5	52.38	59.79	58.89	60.61	72.86	75.47	46.37	53.25	48.54	54.42	56.36	59.68
6	51.59	58.78	57.63	59.01	69.67	73.18	44.14	50.14	46.13	51.28	54.78	57.75
7	52.96	57.61	57.80	59.36	70.27	71.39	45.56	47.76	45.64	48.24	51.87	55.19
8	53.50	55.87	55.97	56.12	64.03	68.26	43.31	44.57	41.29	45.11	49.41	51.48
9	51.16	56.45	57.26	57.45	65.39	69.87	38.76	40.36	39.56	42.56	48.35	49.26
10	51.83	55.61	56.85	57.07	64.10	67.85	34.58	35.49	37.45	40.76	45.61	47.25
Avg.	53.06	58.22	57.57	59.01	68.27	71.35	43.29	48.65	44.84	50.47	54.57	56.92

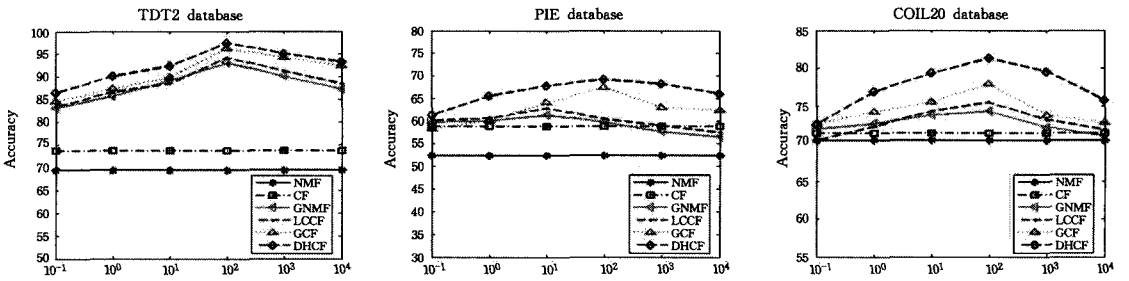
表 3 COIL20 数据库上的聚类实验结果

k	Accuracy/%						Normalized Mutual Information/%					
	NMF	GNMF	CF	LCCF	GCF	DHCF	NMF	GNMF	CF	LCCF	GCF	DHCF
2	89.84	90.11	89.72	90.74	92.48	94.76	71.25	73.78	71.13	74.51	80.40	85.62
3	77.80	83.23	79.34	84.22	85.36	89.23	63.42	69.41	63.21	68.69	76.35	82.37
4	73.01	77.35	73.04	78.14	82.69	86.68	67.87	68.96	66.38	70.63	77.43	82.63
5	70.36	75.19	71.33	74.46	79.23	84.15	66.07	67.83	67.67	72.22	78.56	83.24
6	65.20	73.45	75.21	79.59	82.90	87.87	68.34	69.03	65.33	68.81	74.89	79.42
7	64.64	71.18	63.85	70.08	73.62	81.32	70.14	70.96	66.67	70.57	75.31	80.17
8	65.16	71.39	64.64	71.64	75.51	79.69	70.40	71.34	67.28	70.67	76.45	81.27
9	64.87	68.74	62.86	67.87	70.02	75.41	71.65	72.28	66.40	69.86	72.71	78.55
10	65.37	66.52	62.15	65.71	68.44	70.59	71.89	71.57	66.27	68.69	70.63	73.25
Avg.	70.69	75.24	71.35	75.83	78.92	83.30	69.00	70.57	66.70	70.52	75.86	80.72

4.2 参数选择

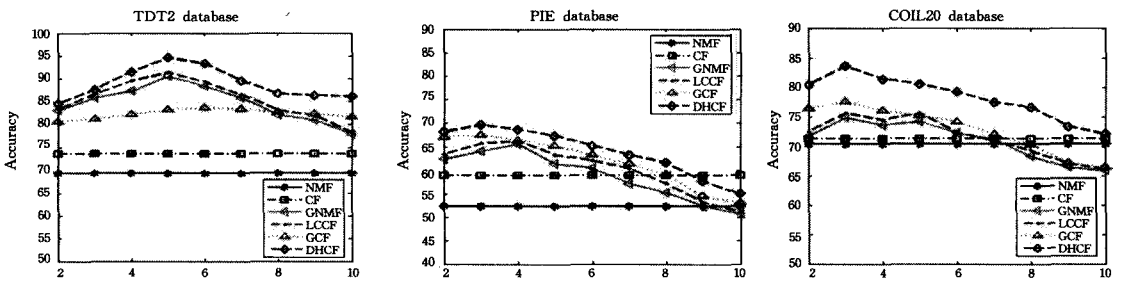
提出的 DHCF 算法主要有 3 个参数:构造超边时所需近邻点数 p 和两个正则化参数 α 和 β。为简便起见,设定两个正则化参数为相同的数值,即 $\alpha=\beta$。故下面讨论 DHCF 算法关于两个参数的稳定性,即给出算法参数设定值与算法的聚类准确率(Accuracy)之间的变化关系。当讨论正则化参

数 α 时，设定创建超边时所选择的近邻点数 p 等于 5。类似地，讨论 p 时，设定正则化参数 α 等于 100。为了同时能与 GNMF，LCCF 和 GCF 算法进行比较，实验中也设定 GNMF，LCCF 和 GCF 算法中的图正则参数与近邻点数的取值与本文相同。分别在文本数据库 TDT2 以及图像数据库 PIE 和 COIL20 中进行了比较实验；同时，NMF 和 CF 算法作为参考数值，将其一并给出。实验结果如图 1 和图 2 所示。



(a) regularization parameters $\alpha=\beta$　(b) regularization parameters $\alpha=\beta$　(c) regularization parameters $\alpha=\beta$

图 1　各种算法的聚类准确率随正则化参数改变的变化情况



(a) Number of neighborhoods p　(b) Number of neighborhoods p　(c) Number of neighborhoods p

图 2　各种算法的聚类准确率随构建超边顶点数改变的变化情况

由图 1 和图 2 可以得到如下的结论：

(1)提出的 DHCF 算法对于两个正则参数而言是非常稳定的。当两个正则化参数在 1 至 1000 的范围内取值时，算法能够取得不错的聚类效果。同时，DHCF 算法本身所得的结果要优于其他算法。

(2)提出的 DHCF 算法的聚类准确率随着最近邻数取值增大到一定程度后而降低，这是由于过大的最近邻数生成的图不能准确地反映样本间固有的几何结构。

结束语　本文提出了一种基于对偶超图正则化的概念分解算法，可同时利用数据流形和特征流形的几何结构信息，通过分别在数据空间和特征属性空间中构建无向加权的拉普拉斯超图正则项，分别反映了数据流形和特征流形的多元几何结构信息，弥补了传统图模型只能表达数据间成对关系的缺陷；本文还给出了 GHCF 算法的目标函数及迭代更新公式，证明了算法的收敛性。大量的实验结果表明提出的算法比已有相关算法在数据表示方面的性能更好。

参 考 文 献

[1] YI Y G, SHI Y J, ZHANG H, et al. Label propagation based semi-supervised nonnegative matrix factorization for feature extraction[J]. Neurocomputing, 2015, 149(PB): 1021-1037.

[2] WANG M M, ZUO W L, WANG Y. A Multidimensional Personality Traits Recognition Model Based on Weighted Nonnegative Matrix Factorization[J]. Chinese Journal of Computers, 2016, 39(38): 1-17. (in Chinese)
王萌萌，左万利，王英. 一种基于加权非负矩阵分解的多维用户人格特质识别算法[J]. 计算机学报，2016，39(38)：1-17.

[3] HE C B, TANG Y, YANG A T, et al. Large-scale topic community mining based on distributed nonnegative matrix factorization[J]. SCIENTIA SINICA Informationis, 2016, 46(6): 714-728. (in Chinese)
贺超波，汤庸，杨阿祧，等. 基于分布式非负矩阵分解的大规模主题社区挖掘[J]. 中国科学：信息科学，2016，46(6)：714-728.

[4] LEE D D, SEUNG H S. Learning the parts of objects by nonnegative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.

[5] XU W, GONG Y H. Document clustering by concept factorization[C]// International Conference on Research and Development in Information Retrieval. Sheffield, UK, 2004: 202-209.

[6] SHANG F H, JIAO L C, WANG F. Graph dual regularization non-negative matrix factorization for co-clustering[J]. Pattern Recognition, 2012, 45(6): 2237-2250.

[7] TENENBAUM J B, SILVA V D, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.

[8] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.

[9] BELKIN M, NIYOGI P. Laplacian Eigenmaps and spectral techniques for embedding and clustering[J]. Advances in Neural Information Processing Systems, 2001, 14(6): 585-591.

[10] HU X K, SUN F M, LI H J. Constrained Nonnegative Matrix Factorization with Sparseness for Image Representation[J]. Computer Science, 2015, 42(7): 280-284. (in Chinese)
胡学考，孙福明，李豪杰. 基于稀疏约束的半监督非负矩阵分解算法[J]. 计算机科学，2015，42(7)：280-284.

[11] FANG W T, MA P, CHENG Z B, et al. 2-dimensional Projective

Non-negative Matrix Factorization and Its Application to Face Recognition[J]. Acta Automatica Sinica, 2012, 38(9): 1503-1512. (in Chinese)
方蔚涛,马鹏,成正斌,等. 二维投影非负矩阵分解算法及其在人脸识别中的应用[J]. 自动化学报,2012,38(9):1503-1512.

[12] CAI D, HE X F, HAN J W, et al. Graph regularization non-negative matrix factorization for data representation[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560.

[13] WANG J Y, BENSMAIL H, GAO X. Multiple graph regularized nonnegative matrix factorization[J]. Pattern Recognition, 2013, 46(10): 2840-2847.

[14] GUAN N Y, TAO D C, LUO Z G, et al. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent[J]. IEEE Trans on Image Processing, 2011, 20(7): 2030-2048.

[15] CAI D, HE X F, HAN J W. Locally consistent concept factorization for document clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(6): 902-913.

[16] YE J, JIN Z. Dual-graph regularized concept factorization for clustering[J]. Neurocomputing, 2014, 138(3): 120-130.

[17] ZHOU D Y, HUANG J Y, SCHOLKOPF B. Learning with hypergraphs: clustering, classification and embedding[C]// Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 2006: 1601-1608.

[18] HUANG Y C, LIU Q S, ZHANG S T, et al. Image retrieval via probabilistic hypergraph ranking[C]// Proceedings of the International Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 3376-3383.

[19] WEI B, CHENG M, WANG C, et al. Combinative hypergraph learning for semi-supervised image classification [J]. Neurocomputing, 2015, 153: 217-277

[20] JIN T, YU J, YOU J, et al. Low-rank matrix factorization with multiple hypergraph regularizer [J]. Pattern Recognition, 2015, 48(3): 1011-1022.

[21] ZENG K, YU J, LI C, et al. Image clustering by hyper-graph regularized non-negative matrix factorization [J]. Neurocomputing, 2014, 138(11): 209-217.

(上接第 292 页)

足视频颜色校正系统平台的实时性要求,能够应用到视频实时处理平台上。

参 考 文 献

[1] LI J F, FANG J L, DAI W Z. No-reference image quality assessment based on color perception[J]. Chinese Journal of Scientific Instrument, 2015, 36(2): 339-350. (in Chinese)
李俊锋,方建良,戴文战. 基于色彩感知的无参考图像质量评价[J]. 仪器仪表学报,2015,36(2):339-350.

[2] MOREL J M, PETRO A B, SBERT C. A PDE formalization of Retinex theory[J]. IEEE Transactions on Image Process, 2010, 22(12): 2825-2837.

[3] WANG R G, ZHANG X, ZHANG X L, et al. A Novel Adaptive Retinex Algorithm for Image Enhancement[J]. Acta Electronica Sinica, 2010, 38(12): 2933-2936. (in Chinese)
王荣贵,张璇,张新龙,等. 一种新型自适应 Retinex 图像增强方法研究[J]. 电子学报,2010,38(12):2933-2936.

[4] FANG S, YANG J R, CAO Y, et al. Local multi-scale Retinex algorithm based on guided image filtering[J]. Journal of Image and Graphics, 2012, 17(7): 748-755.

[5] YANG W T, YANG R G, FANG S, et al. Variable Filter Retinex Algorithm for Foggy Image Ehancement[J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(6): 965-971. (in Chinese)
杨万挺,王荣贵,方帅,等. 滤波器可变的 Retinex 雾天图像增强算法[J]. 计算机辅助设计与图形学学报,2010,22(6):965-971.

[6] PREMKUMAR S, PARTHASARATHI K A. An efficient approach for colour image enhancement using discrete shearlet transform[C]// International Conference on Current Trends in Engineering and Technology. 2014: 363-366.

[7] CHENG F J, DU X J, MA L, et al. Low-light Image Enhancement Based on Retinex[J]. Video Engineering, 2013, 37(15): 4-10. (in Chinese)
程芳瑾,杜晓骏,马丽,等. 基于 Retinex 的低照度图像增强[J]. 电视技术,2013,37(15):4-10.

[8] YIN J C, LI H B, DU J P, et al. Low Illumination Image Retinex Enhancement Algorithm Based on Guided Filtering[C]// Proceedings of Communications in Computer and Information Science. 2014: 639-644.

[9] GUO Y G, GE Q P, GUO N. Colour Correction Based on White Balance[J]. Computer Engineering and Applications, 2005, 41(20): 56-59. (in Chinese)
郭永刚,葛庆平,郭楠. 利用白平衡进行偏色图像的颜色校正[J]. 计算机工程与应用,2005,41(20):56-59.

[10] FU H, HUANG X Y. Location Method of Leaf Image Specular Reflection Regions Based on Retinex[J]. Computer Engineering, 2010, 36(24): 197-199. (in Chinese)
付慧,黄心渊. 基于 Retinex 的树叶图像镜面反射区定位方法[J]. 计算机工程,2010,36(24):197-199.

[11] YUAN Y, TAN Y H, SUN H X, et al. Design and implementation of SCA based ON ZedBoard[J]. Application of Electronic Technique, 2015, 41(11): 31-33. (in Chinese)
袁扬,谭月辉,孙慧贤,等. 基于 ZedBoard 的 SCA 架构的设计与实现[J]. 电子技术应用,2015,41(11):31-33.

[12] GONG Y H, WEI D B, QIAO L Y, et al. Developed of NAND Flash storage system based on Zynq[J]. Electronic Measurement Technology, 2014, 37(12): 53-57. (in Chinese)
龚有华,魏德宝,乔立岩,等. 基于 Zynq 的 NAND Flash 存储系统研制[J]. 电子测量技术,2014,37(12):53-57.

[13] WANG Y Y, SU B H, QIU W S. Super-resolution video restoration system based on ZedBoard[J]. Journal of Applied Optics, 2015, 36(4): 537-542. (in Chinese)
王源圆,苏秉华,邱文胜. 基于 ZedBoard 的超分辨率视频复原系统[J]. 应用光学,2015,36(4):537-542.

[14] LU Q Q, XI D D. Transplantation of embedded Linux system [J]. Foreign Electronic Measurement Technology, 2014, 33(12): 78-81. (in Chinese)
路青起,席丹丹. 嵌入式 Linux 系统移植[J]. 国外电子测量技术,2014,33(12):78-81.