

文章编号:1004-9037(2014)01-0011-08

一种潜在信息约束的非负矩阵分解方法

高新波 王 笛 王秀美

(西安电子科技大学电子工程学院,西安,710071)

摘要:传统的非负矩阵分解方法没有充分利用数据间的内在相似性,从而影响了算法的性能。为此,本文提出一种潜在信息约束的非负矩阵分解方法。该方法首先利用迭代最近邻方法挖掘原始数据的潜在信息,然后利用潜在信息构造数据之间的相似图,最后将相似图作为约束项求得非负矩阵的最优分解。相似图的约束使得非负矩阵分解在降维过程中保持了原始数据之间的相似性关系,进而提高了非负矩阵分解的判别能力。图像聚类实验结果表明了该方法的有效性。

关键词:数据降维;非负矩阵分解;潜在信息;相似图;迭代最近邻

中图分类号:TP391

文献标志码:A

Potential Information Restrained Nonnegative Matrix Factorization

Gao Xinbo, Wang Di, Wang Xiumei

(School of Electronic Engineering, Xidian University, Xi'an, 710071, China)

Abstract: Traditional NMF method does not fully utilize the internal similarity among original data, thus the performance of dimensionality reduction is limited. To this end, a new nonnegative matrix factorization algorithm restrained by the regularization of potential information is proposed. Firstly, the potential information is mined via the iterative nearest neighbor. Then the potential information is utilized to construct similarity graph of data set. Finally, the similarity graph is incorporated as a regularization term to preserve the relationship between original data in the decomposition process of nonnegative matrix. The regularization term keeps the similarity between the original data in the process of dimensionality reduction, which can improve the discriminant ability of nonnegative matrix factorization algorithm. Thorough experiments on standard image databases show the superior performance of the proposed method.

Key words: dimensionality reduction; nonnegative matrix factorization; potential information; similarity graph; iterative nearest neighbor

引 言

大数据给机器学习、数据挖掘和模式识别带来了新的机遇与挑战。降维分析能有效避免数据维数灾难问题,并且可以在保留高维数据的大部分内在信息的同时将高维数据映射到低维空间,是挖掘海量、高维数据中有效信息的重要手段。传统的降维方法允许降维后的数据存在负值,如局部线性嵌入(Locally linear embedding, LLE)^[1],等距映射

(Isometric feature mapping, Isomap)^[2],局部保持投影(Locality preserving projections, LPP)^[3],拉普拉斯特征映射(Laplacian eigenmaps, LE)^[4]。但是,实际应用中很多数据具有明确的物理意义,要求非负数值,这种特性被称为“非负性”,如数字图像的灰度值、文本文档中关键词出现的频率等。面对海量高维非负数据,如何针对其非负特性寻找有效的降维方法成为一个亟待解决的重要问题。非负矩阵分解(Nonnegative matrix factorization, NMF)是解决这一问题的经典方法,它将原始非负

数据分解为非负基矩阵和非负系数矩阵,使得重构数据是由基矩阵以累加的方式叠加而成,符合人类思维中“局部构成整体”的观念。非负矩阵分解在达到维数约简的同时还能得到非负数值结果,具有较好的可解释性,成为目前流行的数据处理与分析方法,广泛地应用于图像处理^[5]、文本聚类^[6]、推荐系统^[7]、视频跟踪^[8]、特征提取^[9]、语音信号处理^[10]等方面。

经典的非负矩阵分解方法提出以后,国内外学者从不同角度提出了改进方法。为了增强非负矩阵分解基矩阵的局部特性,Li等人通过对非负矩阵分解的目标函数增加局部约束项提出了局部非负矩阵分解方法(Local NMF, LNMF)^[11]。为了增强非负矩阵分解的稀疏性,Hoyer对非负矩阵分解的目标函数增加了稀疏约束项,提出稀疏非负矩阵分解方法(NMF with sparseness constraints, SNMF)^[12]。Cai等人提出图正则化非负矩阵分解方法(Graph regularized NMF, GNMF),该方法根据原始数据的几何结构构造近邻图,然后将近邻图作为约束项添加到非负矩阵分解的目标函数中,从而使得非负矩阵分解在降维的同时保持原始数据的几何结构关系^[13]。Zhang等人对图正则化非负矩阵分解的目标函数添加正交约束项使所有基向量正交,从而使得图正则化非负矩阵分解有局部最优的解析解^[14]。Wang等人利用不同类型的近邻图的线性组合来代替图正则化非负矩阵分解中的单一近邻图,更好地保持了数据的几何结构信息^[15]。上述算法均取得了较好的聚类 and 识别效果,它们的共同特点是根据特定目标,在非负矩阵分解算法的目标函数中增加约束项来改善非负矩阵分解的性能。

上述基于图正则化的非负矩阵分解方法均利用近邻图来保持数据的几何结构。传统近邻图是通过 k 近邻方法来构造数据的相似性关系,然而 k 近邻方法虽然速度快,但是不能保证较小的重构误差,会对非负矩阵分解精确性造成影响,进而降低非负矩阵的判别性能。迭代最近邻方法^[16]可以得到数据之间互相表示的重构系数,且重构系数具有稀疏性,并且可以保证较小的重构误差。本文采用迭代最近邻方法得到数据之间互相表示的重构系数,其大小反映了数据之间关系的强弱,称之为数据的潜在信息。利用潜在信息构造反映数据之间真实关系的相似图,基于此,本文提出了一种潜在信息约束的非负矩阵分解算法。该算法利用挖掘

出的潜在信息构建相似图,根据相似图构造约束项,并将约束项添加到非负矩阵分解的目标函数中,从而使得非负矩阵分解在降维的过程中保持原始数据之间的相似性关系。

1 非负矩阵分解

对于非负数据矩阵 $\mathbf{X}=[\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbf{R}^{M \times N}$,其中列向量 $\mathbf{x}_i (i=1, \dots, N)$ 表示数据样本点,非负矩阵分解的目的是将非负矩阵 \mathbf{X} 近似分解为两个非负矩阵 $\mathbf{U} \in \mathbf{R}^{M \times K}$ 和 $\mathbf{V} \in \mathbf{R}^{N \times K}$ 乘积的形式,即

$$\mathbf{X} \approx \mathbf{UV}^T \quad (1)$$

其中,矩阵 \mathbf{U} 中的所有列向量可以认为是对数据矩阵 \mathbf{X} 进行线性逼近的一组基,而 \mathbf{V} 可以认为是数据矩阵 \mathbf{X} 在新基 \mathbf{U} 下的投影系数。通常情况下, $K \ll M, N$,即非负矩阵分解将原始的 M 维数据约减成了 K 维数据,起到了降维和压缩的作用。

在非负矩阵分解中,寻找一个近似的分解过程是最重要的核心问题,这就需要定义一个目标函数,用优化的方式来逼近矩阵的分解。一般地,利用两个矩阵之间的欧氏距离来衡量逼近效果,目标函数为^[17]

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} f(\mathbf{U}, \mathbf{V}) &= \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \\ \text{s. t. } \mathbf{U}, \mathbf{V} &\geq 0 \end{aligned} \quad (2)$$

由于目标函数(2)不是凸函数,要找到一个全局最优解是非常困难的。因此, Lee等人利用乘性迭代算法来寻求目标函数的局部最优解^[18],其迭代规则为

$$\mathbf{U}_{ij} \leftarrow \mathbf{U}_{ij} \frac{(\mathbf{XV})_{ij}}{(\mathbf{UV}^T\mathbf{V})_{ij}} \quad (3)$$

$$\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij} \frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{VU}^T\mathbf{U})_{ij}} \quad (4)$$

重复迭代式(3,4)可以逐渐减小目标函数(2),从而使得目标函数收敛于某个局部最优分解。

2 潜在信息约束的非负矩阵分解方法

高维空间中的数据间通常存在一定的内在关系,可以利用权值大小来描述数据之间关系的强弱,并在降维的过程中保持这种权值关系,从而达到降维后的数据能够保持原始数据之间几何结构关系的目的^[19-20]。在非负矩阵分解中,通过构造样本之间的近邻图作为约束条件,以此来保持非负矩阵分解前后样本之间的几何结构关系。然而传统近邻图是通过 k 近邻方法来构造,由于 k 近邻方法

对近邻数目的强制性约束,会产生较大的截断误差,从而对非负矩阵分解造成较大的重构误差,影响非负矩阵分解的性能。

本文通过学习的方法,挖掘数据之间的潜在信息,并根据数据之间的潜在信息构建一个能够反映数据点之间真实关系的相似图,进而合理利用该相似图改善非负矩阵分解的性能,提高非负矩阵分解的辨别能力。

2.1 潜在信息挖掘

高维数据之间通常会存在一定的内在关系,在降维的过程中保持数据之间的相似性关系至关重要。迭代最近邻方法^[16](Iterative nearest neighbors, INN)可以得到数据之间的相似性程度。对于给定的字典 $D=[d_1, \dots, d_N] \in \mathbf{R}^{M \times N}$ 和一个测试数据 $q \in \mathbf{R}^M$, 迭代最近邻的优化目标函数为

$$\arg \min_{\{s_i\}_{i=1}^K} \left\{ q - \sum_{i=1}^K \frac{\gamma}{(1+\gamma)^i} s_i \right\} \quad (5)$$

其中 K 为迭代次数, $s_i \in D, \gamma \in (0, 1)$ 。通过迭代,可以由字典 D 重建测试数据 q , 即

$$q \approx c_1 s_1 + c_2 s_2 + \dots + c_K s_K \quad (6)$$

其中重构系数 $c_i = \frac{\gamma}{(1+\gamma)^i}, i=1, \dots, K$ 。由于 γ 非负,所以 c_i 非负,且满足

$$\lim_{K \rightarrow \infty} \left\{ \sum_{i=1}^K c_i \right\} = 1 \quad (7)$$

式中: c_i 越大, s_i 对重构 q 的贡献越大,则说明 s_i 与 q 的关系越强。迭代最近邻算法过程如表 1 所示。

表 1 INN 算法

Table 1 INN Algorithm

输入:	$D \in \mathbf{R}^{M \times N}, q \in \mathbf{R}^M, \gamma \in (0, 1), \epsilon \in \mathbf{R}_+, K \in \mathbf{N}_+$
输出:	$\{s_i\}_1^K, \{c_i\}_1^K$
(1) 初始化:	$k=0, q'=q, \hat{q}=0, \omega_0=\gamma;$
(2) 在 D 中查找测试数据 q' 的最近邻样本 s_k :	$k=k+1, s_k = NN(q'), c_k = \frac{\omega_k - 1}{1 + \gamma};$
(3) 更新重构项 \hat{q} 和误差项 r :	$\hat{q} = \hat{q} + c_k s_k, r = \ q - \hat{q}\ _2;$
(4) 如果 $k < K$ 并且 $r > \epsilon$, 返回(2)。	

图 1 为 INN 算法在 Yale 人脸数据库的实验结果。本文随机在 Yale 数据库中选取一幅人脸图像作为测试图像,即为左上角图像,利用 INN 算法在由其余图像组成的字典中对测试图像进行重构,得到重构系数 c 。从图中可以看出,测试图像可以由第三行的最近邻样本图像以及对应的重构系数进行重构得到,其中重构图像为右下角图像。在第三行的最近邻样本中,用来重构测试图像的最近邻图像和测试图像均属于同一个人。因此,可以利用得到的重构系数来表示测试数据和字典中数据之间的潜在关系,且重构数据对应的系数越大则代表其与测试数据的关系越紧密。

由于 INN 算法可以得到测试数据和字典中数据的潜在关系,可以通过这种潜在关系表示数据之间的相似性,进而构建数据的相似矩阵。对于数据 $x_i \in X=[x_1, \dots, x_N]$,将其余数据组成的字典 $D_i=[x_1, \dots, x_{i-1}, \mathbf{0}, x_{i+1}, \dots, x_N]$,利用迭代最近邻求解 x_i 在字典 D_i 中的重构的系数 C_i















迭代次数	1	2	3	4	5	6
测试数据 q'						
最近邻样本 s						
重构系数 c	0.2	0.16	0.128	0.102 4	0.081 92	0.065 536
重构项 \hat{q}						

图 1 迭代最近邻算法在 Yale 人脸数据库的实验结果

Fig. 1 Experimental result of INN on the Yale face data set

$$\mathbf{x}_i = \mathbf{D}_i \mathbf{C}_i \quad (8)$$

分别对训练数据 \mathbf{X} 中其他样本做相同的处理,最终得到训练集中每个数据点与其他数据点之间的潜在关系,并用矩阵 $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_N]$ 表示训练数据间的潜在关系矩阵。

2.2 潜在信息约束的非负矩阵分解方法

为了改善非负矩阵分解的性能,利用上述得到的潜在关系矩阵构造数据的相似图,并根据相似图构造非负矩阵分解目标函数中的约束项,从而使得降维后的数据可以很好地保留原始数据之间的相似关系。

对于数据 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbf{R}^{M \times N}$, 构造 k 近邻图 $\mathbf{W} = \{w_{ij}\}_{N \times N}$, 利用约束矩阵 \mathbf{C} 调整近邻图 \mathbf{W} , 得到数据之间的相似图 $\tilde{\mathbf{W}} = \{\tilde{w}_{ij}\}_{N \times N}$

$$\tilde{w}_{ij} = c_{ij} + w_{ij} - c_{ij}w_{ij} \quad (9)$$

为了使 $\tilde{\mathbf{W}}$ 对称, 令 $\tilde{\mathbf{W}} = (\tilde{\mathbf{W}} + \tilde{\mathbf{W}}^T) / 2$ 。

根据以上规则构造的相似图 $\tilde{\mathbf{W}}$ 有如下性质:

- (1) $\tilde{\mathbf{W}}$ 是对称的, 且 $\tilde{w}_{ij} \in [0, 1]$;
- (2) $\tilde{w}_{ij} > w_{ij}$, 如果 $c_{ij} > 0$;
- (3) $\tilde{w}_{ij} > c_{ij}$, 如果 $c_{ij} > 0$ 。

性质(2)说明当 c_{ij} 大于 0, 即 x_i 和 x_j 具有较高的相似性, 相似图 $\tilde{\mathbf{W}}$ 的权值大于近邻图 \mathbf{W} 的权值, 从而更好地体现了数据之间的相似性。性质(3)可以保证, 当 c_{ij} 值较大时, 即使近邻图中权值 w_{ij} 很小, 相似图中 \tilde{w}_{ij} 也会较大。图 2 为相似图构造过程示意图, 图 2(a) 和图 2(c) 为黑色数据点 x_1 的 k 近邻图, 其余点距离 x_1 越近, 其与 x_1 的对应

权值越大, 其中红色点代表点 x_1 的 k 近邻点, 蓝色点代表在 x_1 的近邻图中权值较小而潜在关系矩阵中权值较大的点。假设 x_1 的 k 近邻点对应的潜在关系 c_{1j} 大于 0, 根据性质(2), 在相似图 2(b) 中, x_1 的 k 近邻点较原始 k 近邻图 2(a) 有更强的相似性。根据性质(3), 在相似图 2(d) 中, 潜在关系矩阵中权值较大的蓝色点较原始 k 近邻图 2(c) 有更强的相似性。因此, 根据潜在关系矩阵构造的相似图较 k 近邻图更好地保持了数据之间的相似性。

为了保持数据之间的相似关系, 本文构造如下约束项

$$\varphi(\mathbf{V}) = \frac{1}{2} \sum_{i,j=1}^N \| \mathbf{v}_i - \mathbf{v}_j \|^2 \tilde{w}_{ij} \quad (10)$$

当 c_{ij} 为正时, x_i 和 x_j 相似性较大, 则 x_i 和 x_j 在降维空间中也应该相似, 由于 \tilde{w}_{ij} 较大, 为使 $\varphi(\mathbf{V})$ 较小, 则 \mathbf{v}_i 和 \mathbf{v}_j 也应该相似。这样, 就达到了原始空间中相似性较高的两个数据, 在降维空间中也相似的目的, 从而让样本之间的相似性关系得以保持。因此, 潜在信息约束的非负矩阵分解方法的目标函数为

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} f(\mathbf{U}, \mathbf{V}) &= \| \mathbf{X} - \mathbf{UV}^T \|_F^2 + \lambda \varphi(\mathbf{V}) \\ \text{s. t. } \mathbf{U}, \mathbf{V} &\geq 0 \end{aligned} \quad (11)$$

式中右边第一项是重构项, 用来寻找一个新的低维空间表示高维空间中的样本。第二项是约束项, 使得高维空间中样本间的相似性关系在低维空间中还能够得以保持。 λ 用于平衡两部分在降维过程中的比重。

2.3 迭代准则

为了得到最终的迭代准则, 可以对式(10)进一步简化

$$\begin{aligned} \varphi(\mathbf{V}) &= \frac{1}{2} \sum_{i,j=1}^N \| \mathbf{v}_i - \mathbf{v}_j \|^2 \tilde{w}_{ij} = \sum_{i=1}^N \mathbf{v}_i^T \mathbf{v}_i \tilde{\mathbf{D}}_{ii} - \\ &\sum_{i,j=1}^N \mathbf{v}_i^T \mathbf{v}_j \tilde{w}_{ij} = \text{Tr}(\mathbf{V}^T \tilde{\mathbf{D}} \mathbf{V}) - \text{Tr}(\mathbf{V}^T \tilde{\mathbf{W}} \mathbf{V}) = \\ &\text{Tr}(\mathbf{V}^T \tilde{\mathbf{L}} \mathbf{V}) \end{aligned} \quad (12)$$

其中 $\tilde{\mathbf{D}}$ 是一个对角阵, 第 i 个对角元素为 $\sum_j \tilde{w}_{ij}$, 矩阵 $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$ 。将式(12)代入式(11)可以得到

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} f(\mathbf{U}, \mathbf{V}) &= \text{Tr}((\mathbf{X} - \mathbf{UV}^T)(\mathbf{X} - \mathbf{UV}^T)^T) + \\ &\lambda \text{Tr}(\mathbf{V}^T \tilde{\mathbf{L}} \mathbf{V}) = \text{Tr}(\mathbf{X} \mathbf{X}^T) - 2 \text{Tr}(\mathbf{X} \mathbf{V} \mathbf{U}^T) + \\ &\text{Tr}(\mathbf{U} \mathbf{V}^T \mathbf{V} \mathbf{U}^T) + \lambda \text{Tr}(\mathbf{V}^T \tilde{\mathbf{L}} \mathbf{V}) \end{aligned} \quad (13)$$

由于存在两个约束条件 $\mathbf{U}, \mathbf{V} \geq 0$, 分别定义 Lagrange 乘子 Ψ_{ij} 和 φ_{ij} , 记 $\Psi = [\Psi_{ij}]$ 和 $\Phi = [\varphi_{ij}]$ 分别表示 Lagrange 乘子 Ψ_{ij} 和 φ_{ij} 组成的矩阵。因

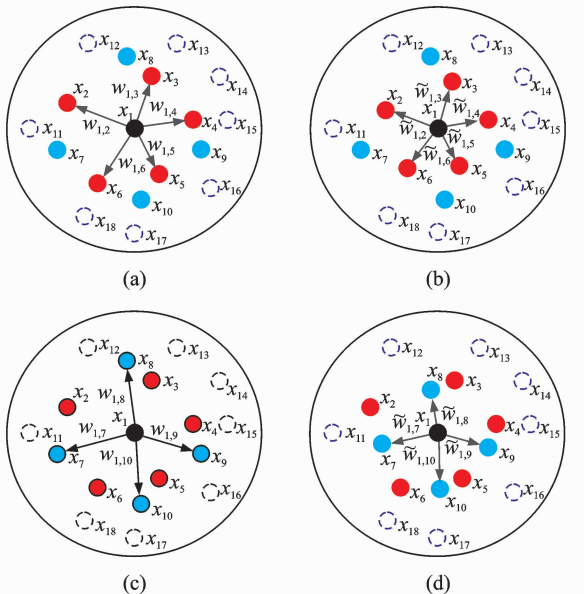


图 2 相似图构造过程示意图

Fig. 1 Schematic diagram of construction process for similarity graph

此,新的目标函数可以写为

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} f(\mathbf{U}, \mathbf{V}) = & \text{Tr}(\mathbf{X}\mathbf{X}^T) - 2\text{Tr}(\mathbf{X}\mathbf{V}\mathbf{U}^T) + \\ & \text{Tr}(\mathbf{U}\mathbf{V}^T\mathbf{V}\mathbf{U}^T) + \lambda\text{Tr}(\mathbf{V}^T\tilde{\mathbf{L}}\mathbf{V}) + \\ & \text{Tr}(\Psi\mathbf{U}^T) + \text{Tr}(\Phi\mathbf{U}^T) \end{aligned} \quad (14)$$

为了得到局部最优解,对式(14)分别对 \mathbf{U} 和 \mathbf{V} 求偏导

$$\frac{\partial f}{\partial \mathbf{U}} = -2\mathbf{X}\mathbf{V} + 2\mathbf{U}\mathbf{V}^T\mathbf{V} + \Psi \quad (15)$$

$$\frac{\partial f}{\partial \mathbf{V}} = -2\mathbf{X}^T\mathbf{U} + 2\mathbf{V}\mathbf{U}^T\mathbf{U} + 2\lambda\tilde{\mathbf{L}}\mathbf{V} + \Phi \quad (16)$$

利用 KKT 条件, $\Psi_{ij}\mathbf{U}_{ij} = 0$ 和 $\Phi_{ij}\mathbf{V}_{ij} = 0$, 进一步化简式(15)和式(16)

$$-(\mathbf{X}\mathbf{V})_{ij}\mathbf{U}_{ij} + (\mathbf{U}\mathbf{V}^T\mathbf{V})_{ij}\mathbf{U}_{ij} = 0 \quad (17)$$

$$\begin{aligned} & -(\mathbf{X}^T\mathbf{U})_{ij}\mathbf{V}_{ij} + (\mathbf{V}\mathbf{U}^T\mathbf{U})_{ij}\mathbf{V}_{ij} + \\ & \lambda(\tilde{\mathbf{L}}\mathbf{V})_{ij}\mathbf{V}_{ij} = 0 \end{aligned} \quad (18)$$

最终可以得到 \mathbf{U} 和 \mathbf{V} 的迭代公式

$$\mathbf{U}_{ij} \leftarrow \mathbf{U}_{ij} \frac{(\mathbf{X}\mathbf{V})_{ij}}{(\mathbf{U}\mathbf{V}^T\mathbf{V})_{ij}} \quad (19)$$

$$\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij} \frac{(\mathbf{X}^T\mathbf{U} + \lambda\tilde{\mathbf{W}}\mathbf{V})_{ij}}{(\mathbf{V}\mathbf{U}^T\mathbf{U} + \lambda\tilde{\mathbf{D}}\mathbf{V})_{ij}} \quad (20)$$

3 实验结果

为了验证本文提出的非负矩阵分解算法的有效性,作者在一些标准图像数据库上进行聚类实验,并和本领域的一些主流算法进行对比分析。

3.1 聚类效果评价准则

本文通过对比每个数据的聚类标签和数据集提供的标签,以准确率和归一化互信息准则来衡量算法的聚类效果。

准确率(Accuracy, AC)准则用来计算聚类算法的正确标签比率,衡量算法的准确性。给定一个数据集,样本数为 n , 样本的真实标签为 r_i 。假定 l_i 是通过聚类得到的样本标签,则准确率可以定义为

$$AC = \frac{\sum_{i=1}^n \delta(r_i, \text{map}(l_i))}{n} \quad (21)$$

其中 $\delta(x, y)$ 当 $x = y$ 时,其值为 1, 否则,其值为 0。 $\text{map}(l_i)$ 是最优投影函数,将聚类得到的标签 l_i 投影为对应的样本标签。本文利用 Kuhn-Munkres 算法^[21]来进行投影。

归一化互信息(Normalized mutual information, MI)准则用来估计两个聚类簇之间的相似性。给定两个聚类簇 C 和 C' , 它们的互信息(Mutual information, $\text{MI}(C, C')$)定义为

$$\text{MI}(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \cdot \log \frac{p(c_i, c'_j)}{p(c_i)p(c'_j)} \quad (22)$$

其中 $p(c_i), p(c'_j)$ 分别表示从数据集中的样本属于聚类簇 C, C' 的概率, $p(c_i, c'_j)$ 表示从数据集中样本同时属于聚类簇 C 和 C' 的概率。 $\text{MI}(C, C')$ 的值介于 0 和 $\max(H(C), H(C'))$ 之间,其中 $H(C)$ 和 $H(C')$ 分别为聚类簇 C 和 C' 的熵。当两个聚类簇 C 和 C' 完全相同时, $\text{MI}(C, C')$ 取得最大值 $\max(H(C), H(C'))$, 当两个聚类簇完全独立时, $\text{MI}(C, C')$ 取得最小值 0。在实验中,利用归一化的互信息 $\widehat{\text{MI}}(C, C')$, 其值介于 0 和 1 之间。

$$\widehat{\text{MI}}(C, C') = \frac{\text{MI}(C, C')}{\max(H(C), H(C'))} \quad (23)$$

3.2 实验参数设置

为验证算法的性能,本实验比较潜在信息约束的非负矩阵分解方法(Information-restrained non-negative matrix factorization, INMF)与现有的三个非负矩阵分解的主流算法 NMF^[17], GNMF^[13], MMF^[14] 在三个图像数据集的聚类效果。表 2 给出了各个数据集的相关信息。

表 2 图像数据库
Table 2 Image Data Sets

数据集	图像总数	维度	类别数
PIE	2 856	1 024	68
COIL20	1 440	1 024	20
AR	3 288	3 600	116

为使对比实验更加合理,本文实验条件参照 GNMF 的中条件进行设置。实验过程如下:

(1) 从每个数据集中任意选择 P 类图像作为数据集 X 。

(2) 对于每一个数据集 X , 通过 NMF, INMF, GNMF, MMF 得到原始数据的低维表示 V , 并设低维空间维度为类别数 P 。

(3) 利用 K-means 算法对不同算法得到的低维表示 V 和原始数据进行聚类。重复运行 25 次, 每次采用不同的初始值, 取 25 次实验中使得 K-means 代价函数最小的结果作为每个算法的最终聚类结果。

(4) 利用不同算法的聚类标签和样本的真实标签进行对比得到其对应的 AC 和 $\widehat{\text{MI}}$ 。

(5) 对于每个 P 值, 重复以上过程 20 次, 取 AC 和 $\widehat{\text{MI}}$ 的平均值作为最终结果。

(6) NMF 是无参数方法。GNMF, MMF 和

INMF 含有约束项比重参数 λ , 均设为 100。这 3 种方法需要构造近邻图 W , 均采用热核形式, 如式

$$W = \left[w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2 / 2t^2) & x_i \in N(x_j) \text{ 或者 } x_j \in N(x_i) \\ 0 & \text{其他} \end{cases} \right] \quad (24)$$

3.3 结果与分析

3.3.1 PIE 人脸数据库

PIE 人脸数据库包括 68 个人的 13 种不同姿态、43 种不同光照和 4 种不同表情下的 41 368 张人脸图像。本文利用 PIE 人脸数据库的一个子集^[13]进行实验, 包括 68 个人随光照条件变化的 42 幅正面人脸图像。每张图像均是 32×32 的灰度图像。

图 3 为随着图像类数 P 的变化, 准确率和归一化互信息的变化情况。从图中可以看到, 利用非负矩阵分解降维后的数据进行聚类的效果 (NMF, GNMF, MMF, INMF) 明显优于直接利用原始数据进行聚类

(24) 所示, 其中 t 设为 0.2, 近邻阈值设为 3。INMF 在潜在信息挖掘时含有参数 γ , 设为 0.25。

始 NMF 算法。INMF 算法由于考虑了原始数据的潜在约束信息, 效果要明显优于其他算法。

3.3.2 COIL20 数据库

COIL20 数据库包含 20 个物体的 1 440 幅图像, 其中每个物体从 72 个不同角度进行拍摄。每张图像均是 32×32 的灰度图像。

图 4 为随着图像类数 P 的变化, 准确率和归一化互信息的变化情况。MMF 虽然降维的过程中保留了原始数据的几何结构关系, 但是为了使目标函数具有解析解, 该算法在目标函数中添加正交约束使得所有基向量正交, 而从图 4 中看到, 这种正交约束反而影响了算法的效果。本文提出的 INMF 算法的效果要明显优于其他算法。

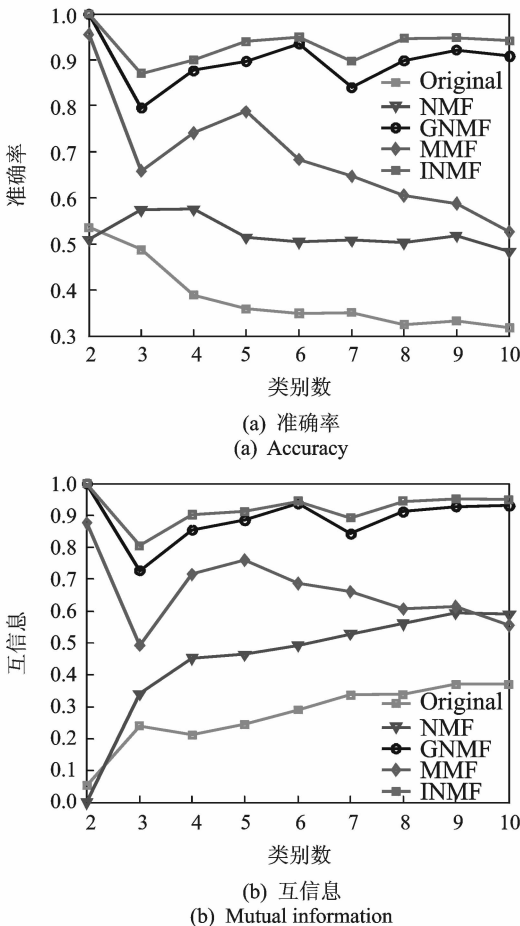


图 3 PIE 人脸数据库实验结果

Fig. 3 Experimental results on the PIE face data set

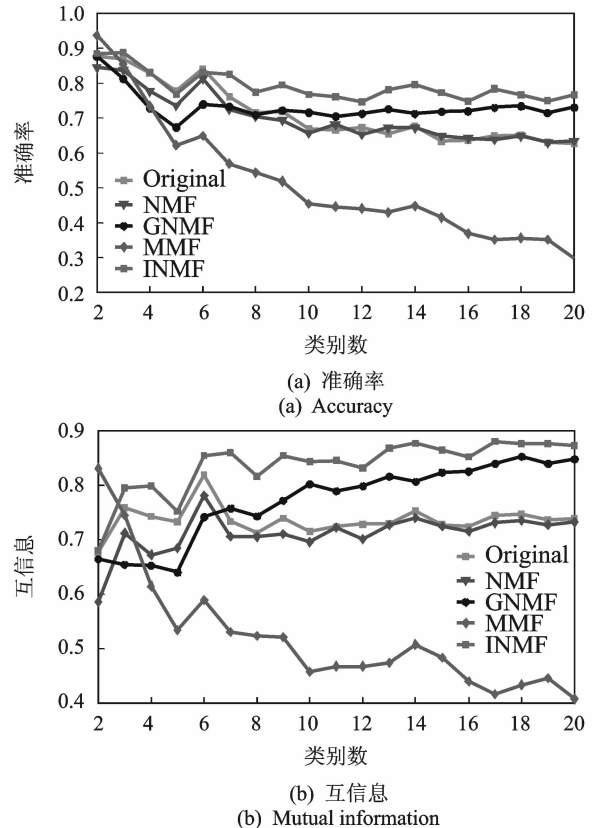


图 4 COIL20 人脸数据库实验结果

Fig. 4 Experimental results on the COIL20 face data set

3.3.3 AR 人脸数据库

AR 人脸数据库包括 116 人不同表情、光照、遮挡和老化的 3 288 幅图像。实验时将每幅图像调整为 60×60 大小。

图 5 为随着图像类数的变化,准确率和归一化互信息的变化情况。图 5(a)中,各个算法的准确率相差不多,但 INMF 整体优于其他算法。图 5(b)中,INMF 在归一化互信息准则下要明显优于其他算法。

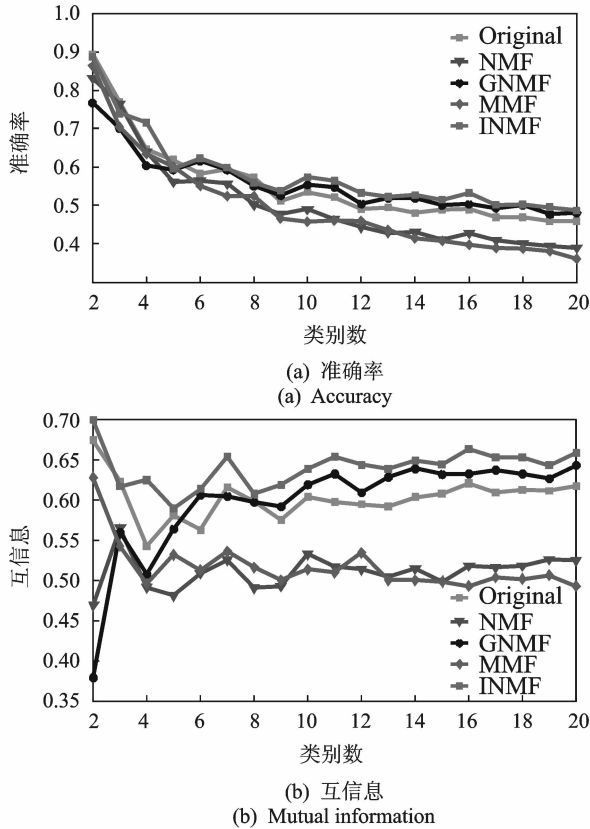


图 5 AR 人脸数据库实验结果

Fig. 5 Experimental results on the AR face data set

4 结束语

本文提出一种潜在信息约束的非负矩阵分解方法,利用迭代最近邻方法构造近邻图,可以在保证最小化重构误差的同时,使得构造的近邻图有较好的稀疏性,从而使非负矩阵分解可以在降维的过程中很好地保持原始数据的相似性关系,进而提高非负矩阵分解方法的判别能力。理论分析和实验结果表明,本文方法具有较好的精确性和有效性。然而本文方法是一种无监督方法,尚未考虑数据的先验信息。当数据有先验信息存在时,可以考虑融合先验信息和潜在信息,从而得到更合理的数据相似图,以便更加有效地提升非负矩阵分解的判别能力。

参考文献:

[1] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000,

290(5500):2323-2326.

- [2] Tenenbaum J, Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000,290(5500):2319-2323.
- [3] He X, Niyogi P. Locality preserving projections [C]//Advances in Neural Information Processing Systems. Whistler, British Columbia, Canada: MIT Press, 2003,16:234-241.
- [4] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]//Advances in Neural Information Processing Systems. Vancouver, British Columbia, Canada: MIT Press, 2001,14:585-591.
- [5] Guillaumet D, Schiele B, Vitria J. Analyzing non-negative matrix factorization for image classification [C]//Proceedings of 16th International Conference on Pattern Recognition. Quebec, Canada: IEEE Computer Society, 2002,2:116-119.
- [6] Shahnaz F, Berry M, Pauca V, et al. Document clustering using nonnegative matrix factorization[J]. Information Processing & Management, 2006, 42 (2):373-386.
- [7] Zhang S, Wang W, Ford J, et al. Learning from incomplete ratings using non-negative matrix factorization[C]//Proceedings of 6th SIAM Conference on Data Mining. Bethesda, MD, USA: SIAM, 2006.
- [8] Bucak S, Günsel B. Video content representation by incremental non-negative matrix factorization [C]//Proceedings International Conference on Image Processing. San Antonio, Texas, USA: IEEE, 2007,2: 113-116.
- [9] Zafeiriou S, Tefas A, Buciu I, et al. Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification[J]. IEEE Transaction on Neural Networks, 2006,17(3): 683-695.
- [10] 孙健,张雄伟,曹铁勇,等.基于卷积非负矩阵分解的语音转换方法[J].数据采与处理,2013,28(2): 141-148.
- Sun Jian, Zhang Xiongwei, Cao Tiejong, et al. Voice conversion based on convolutive nonnegative matrix factorization[J]. Journal of Data Acquisition and Processing, 2013,28(2):141-148.
- [11] Li S, Hou X, Zhang H, et al. Learning spatially localized, parts-based representation [C]//Proceeding Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA: IEEE Computer Society, 2001,1:207-212.
- [12] Hoyer P. Non-negative matrix factorization with

- sparseness constraints[J]. The Journal of Machine Learning Research, 2004,5:1457-1469.
- [13] Cai D, He X, Han J, et al. Graph regularized non-negative matrix factorization for data representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011,33(8):1548-1560.
- [14] Zhang Z, Zhao K. Low rank matrix approximation with manifold regularization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(7):1717-1729.
- [15] Wang J, Bensmail H, Gao X. Multiple graph regularized nonnegative matrix factorization[J]. Pattern Recognition, 2013,46(10):2840-2847.
- [16] Timofte R, Gool L. Iterative nearest neighbors for classification and dimensionality reduction[C]//Proceedings of 16th International Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012,2456-2463.
- [17] Lee D, Seung S. Algorithms for non-negative matrix factorization [C]//Advances in Neural Information Processing Systems. Denver, Colorado, United States: MIT Press, 2000,13:556-562.
- [18] Lee D, Seung S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755):788-791.
- [19] 胡昭华,宋耀良.基于一种连续自编码网络的图像降维和重构[J].数据采集与处理,2010,25(3):318-323.
Hu Zhaohua, Song Yaoliang. Dimensionality reduction and image reconstruction based on continuous autoencoder network[J]. Journal of Data Acquisition and Processing, 2010,25(3):318-323.
- [20] Yan Xuemei, Zhang Limei, Guo Wenbin. Dual-sparsity preserving projection[J]. Transactions of Nanjing Aeronautics and Astronautics, 2012,29(3):284-288.
- [21] Plummer M, Lovasz L. Matching theory[M]. North Holland, 1986.

作者简介:高新波(1972-),男,教授,博士生导师,研究方向:机器学习、模式识别、多媒体内容分析,E-mail:xbgao@mail.xidian.edu.cn;王笛(1989-),女,博士研究生,研究方向:机器学习、数据挖掘;王秀美(1978-),女,副教授,研究方向:机器学习、统计建模。