

改进的局部稀疏表示分类算法及其在人脸识别中的应用

尹贺峰¹ 吴小俊¹ 陈素根^{1,2}

(江南大学物联网工程学院 无锡 214122)¹ (安庆师范学院数学与计算科学学院 安庆 246133)²

摘要 近年来,稀疏表示分类(Sparse Representation Based Classification, SRC)方法在人脸识别中受到越来越多的关注。原始 SRC 方法使用所有的训练样本组成字典矩阵,当训练样本比较多时,稀疏系数的求解会变得非常耗时。为了解决这一问题,提出一种新的局部稀疏表示分类(Local SRC, LSRC)方法。该方法针对每个测试样本,根据测试样本和训练样本稀疏系数之间的相似性来选择部分训练样本,由这些训练样本组成字典,然后在这个字典上对测试样本进行稀疏分解。该方法性能相比于原始 LSRC 方法更稳定。在 ORL、Yale 和 AR 人脸库上的实验结果表明,该方法的效果优于 SRC 和 LSRC。

关键词 稀疏表示分类,局部稀疏表示分类,稀疏系数,相似性,人脸识别

中图分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.8.010

Improved LSRC and its Application in Face Recognition

YIN He-feng¹ WU Xiao-jun¹ CHEN Su-gen^{1,2}

(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)¹

(School of Mathematics & Computational Science, Anqing Normal University, Anqing 246133, China)²

Abstract Recently, sparse representation based classification(SRC) has attracted much attention in face recognition tasks. SRC forms the dictionary by directly using all the training samples. When giving lots of training samples, the speed of the subsequent sparse solver can be very slow. To alleviate this problem, a new local SRC, which is based on the similarities of sparse coefficients of both training samples and test samples, was presented. According to this similarity, a certain number of training samples are selected to form the over-complete dictionary, and then the test sample is decomposed using this dictionary. In contrast to original LSRC, which is based on k NN to choose neighbors of test samples, the proposed approach can steadily achieve better performance. Experimental results obtained on the ORL database, Yale database and AR database indicate that the proposed method is superior to both SRC and LSRC.

Keywords Sparse representation based classification(SRC), Local SRC(LSRC), Sparse coefficients, Similarity, Face recognition

1 引言

人脸识别是计算机视觉和模式识别领域内广泛研究的分类问题之一,传统的人脸识别方法包括 Eigenfaces^[1]、Fisherfaces^[2]、Laplacianfaces^[3]等等,这些方法通常包含两个步骤:特征提取和分类。最近稀疏表示引起了人们越来越多的关注,并被应用于很多计算机视觉和图像处理领域中。尤其是 Wright 等人于 2009 年提出的稀疏表示分类^[4]方法在人脸识别中得到了应用并取得了很好的识别效果。SRC 的思想是给定一个测试样本,首先得到该样本在所有训练样本构成的字典上的稀疏系数,然后利用稀疏系数和每类训练样本重构样本,最后将该测试样本分在最小重构误差所对应的类别中。在 SRC 的框架下,压缩感知的理论表明:特征空间的选择已经不再重要,即使是随机特征也能取得很好的分类结果。关键问题是特征空间的维数足够大并且能正确地计算相应的稀疏表示问题。此外 SRC 对于图像遮挡和像素污损有一定的

鲁棒性。随后又有很多改进的 SRC 方法被提出以提高其性能和鲁棒性,并被应用到其它模式识别问题中。

然而原始的 SRC 算法直接使用所有的训练样本组成字典矩阵,当训练样本数目比较多时,字典的规模比较大,这不利于后面的稀疏分解。为了解决这一问题,Wei 等^[5]提出了局部敏感的字典学习(Locality-sensitive Dictionary Learning)算法,该算法在字典学习过程中加入数据局部约束,取得了比 SRC 好的分类效果;Liu 等^[6]提出了基于 l^2 范数正则化的双线性鉴别字典学习(Bilinear Discriminative Dictionary Learning, BDDL)方法,即使用每类原子来优化每类的重构误差,同时为了使得到的编码系数具有鉴别性,在字典学习过程中引入鉴别项;Yang 等^[7]提出了一种有监督的字典学习方法,该方法基于 Fisher 判别准则,字典中的元素和样本的类别标签对应起来,并且样本稀疏编码后的重构误差也用来进行模式分类。Jiang 等^[8]提出了标签一致 KSVD(Label Consistent K-SVD, LC-KSVD)方法,该方法不仅利用了训练样本的类别

到稿日期:2014-06-10 返修日期:2014-07-26 本文受国家自然科学基金(61373055)资助。

尹贺峰(1988-),男,博士生,主要研究方向为人工智能、模式识别, E-mail: yinhefeng@126.com; 吴小俊(1967-),男,教授,博士生导师,主要研究方向为模式识别、人工智能、计算机视觉; 陈素根(1982-),男,博士生,讲师,主要研究方向为模式识别与智能系统。

标签,还将标签信息和每个字典元素联系起来以增加字典学习过程中稀疏编码的鉴别性。Xu等^[9]提出了两阶段测试样本稀疏表示方法(Two-Phase Test Sample Sparse Representation, TPTSR),第一阶段将测试样本表示为训练样本的线性组合,并利用训练样本就测试样本的表示能力确定测试样本的 M 个邻近样本,第二阶段利用 M 个邻近样本来表示测试样本并利用表示系数进行分类。Zhang等^[10]分析了SRC的机制,指出SRC的性能来源于协同表示而不是范数的稀疏性,基于此,他们将 l^1 范数换成了范 l^2 数,提出了带有正则化最小二乘的协同表示分类(Collaborative Representation based Classification with Regularized Least Square, CRC-RLS)方法。Mi^[11]提出了基于部分训练样本的协同表示分类方法,该方法首先使用多轮细化过程从训练样本中选出部分样本,然后使用协同表示分类器对测试样本进行分类。现有的字典学习算法把降维和字典学习分开进行,这不能充分利用训练样本中的鉴别信息,为了解决这个问题,Feng等^[12]提出了联合鉴别降维和字典学习的算法(Joint Discriminative Dimensionality Reduction and Dictionary Learning, JD-DRDL),JDDRDL同时得到投影矩阵和字典用于人脸表示。此外,Liu等^[13]提出了流形上的字典学习方法,充分考虑了图像的内在几何结构。

与以上改进的稀疏表示分类方法不同,我们关注于SRC中稀疏系数的快速求解。SRC使用所有的训练样本构成字典矩阵,当训练样本数目比较多时,稀疏系数的求解会变得比较困难。文献^[14]提出一种局部稀疏表示分类方法(Local Sparse Representation based Classification, LSRC),此方法利用 k NN准则从训练样本中选择与当前测试样本邻近的部分样本组成字典矩阵,这样字典的规模会降低,测试样本的稀疏分解过程会加快。当样本的维数比较高时, k NN的性能会下降。受稀疏系数相似性可以反映样本之间相似性的启发,根据测试样本和训练样本稀疏系数之间的相似性来选择部分训练样本,由此选择出部分训练样本组成字典,然后在这个字典上对测试样本进行稀疏分解。这样可以在一定程度上避免过多选择其它类别样本,从而提高测试样本的分类准确率。

本文第2节简要介绍相关方法,包括SRC和局部稀疏表示分类方法;第3节提出基于稀疏系数相似性的LSRC算法;第4节是实验验证和分析;最后总结全文。

2 相关工作

2.1 稀疏表示分类算法

稀疏表示分类算法的思想如下^[4]:

假设第 i 类有足够的训练样本,样本图像大小为 $w \times h$,第 i 类有 n_i 个训练样本,把这 n_i 个样本数据均变成列向量,组成矩阵 A_i ,即 $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$, $m = w \times h$ 。给定一幅新的第 i 类的图像 $y \in \mathbb{R}^m$,则 y 可以由第 i 类的所有训练样本近似线性表示:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \dots + \alpha_{i,n_i} v_{i,n_i} \quad (1)$$

其中, $\alpha_{i,j}$ 为标量, $j=1,2,\dots,n_i$ 。

由于在识别之前 y 的类别未知,所以定义一个由所有类别的训练样本组成的矩阵 A ,假设训练样本有 C 类,一共 n 个,即 $\sum_{i=1}^C n_i = n$ 。则

$$A = [A_1, A_2, \dots, A_C] \quad (2)$$

如果用 A 来表示 y ,那么式(1)可以重新写为:

$$y = Ax_0 \in \mathbb{R}^m \quad (3)$$

$x_0 = [0, \dots, 0, \dots, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, \dots, 0, \dots, 0]^T \in \mathbb{R}^n$ 为系数向量,即与第 i 类对应的系数非零,而其它系数均为零。

在稀疏表示中,通常求解最小化 l^1 范数问题:

$$(l^1)x_1 = \arg \min_x \|x\|_1, \text{subject to } Ax = y \quad (4)$$

当图像受到小的噪声影响时,上式可以修改为:

$$(l^1_\epsilon)x_1 = \arg \min_x \|x\|_1, \text{subject to } \|Ax - y\|_2 \leq \epsilon \quad (5)$$

其中, $\epsilon > 0$ 是误差容忍度。

式(5)可以转化为带 l^1 范数约束的最小二乘问题:

$$x_1 = \arg \min_x \|Ax - y\|_2^2 + \lambda \|x\|_1 \quad (6)$$

其中, $\lambda > 0$ 是正则化参数,用来平衡重构误差和稀疏性。当通过求解式(5)或式(6)得到稀疏系数 \hat{x}_1 时,基于重构样本和测试样本之间的残差来对测试样本进行分类。对于第 i 类样本,假设 $\delta_i: \mathbb{R}^m \rightarrow \mathbb{R}^n$ 为特征函数,它选择与第 i 类相关的系数,即对于 $x \in \mathbb{R}^n$, $\delta_i(x) \in \mathbb{R}^n$ 是一个新的向量,这个向量中的非零元素对应 x 中与第 i 类相关的元素,而对应其它类的系数为0。使用 $\delta_i(\hat{x}_1)$ 和由训练样本组成的矩阵 A ,可以得到重构样本。然后就可以把测试样本分在最小重构误差对应的类别中:

$$\min_i r_i(y) = \|y - A\delta_i(\hat{x}_1)\|_2 \quad (7)$$

稀疏表示分类(SRC)算法的流程如下:

- (1)输入:由 C 类训练样本组成的矩阵 $A = [A_1, A_2, \dots, A_C] \in \mathbb{R}^{m \times n}$,测试样本 $y \in \mathbb{R}^m$ (可选参数:容忍误差 $\epsilon > 0$);
- (2)将 y 和 A 的列向量单位化;
- (3)求解式(5)或式(6);
- (4)计算 y 和重构样本的残差 $r_i(y) = \|y - A\delta_i(\hat{x}_1)\|_2$, $i=1,2,\dots,C$;
- (5)输出:测试样本的类别 $\text{identity}(y) = \arg \min_i r_i(y)$ 。

2.2 局部稀疏表示(LSRC)

SRC算法中,在理论上,如果能找到稀疏解,那么测试样本在字典矩阵上的稀疏分解应当是非常快速的,然而在实际应用中,计算稀疏系数是比较耗时的,一方面是由于现存的稀疏表示算法很难在有限时间内找到理想中的稀疏解;另一方面随着训练样本数的增多,相应的字典矩阵会很大,这同样也不利于测试样本的稀疏分解。

文献^[14]提出使用 k NN近邻准则来寻找和测试样本邻近的训练样本,由这些训练样本构成当前测试样本的字典矩阵,然后在这个字典上对测试样本进行稀疏分解。由于SRC方法计算出的稀疏系数有很多接近0的值,这一方面会增加算法的计算时间,另一方面会降低SRC的性能。使用部分训练样本的好处是会加快测试样本的稀疏分解过程,同时没有被选择的训练样本对应的系数刚好就是0,这样可以提高SRC的分类准确性。

LSRC求解的优化问题为:

$$\hat{x}_1 = \arg \min_x \|A_{N(y)}x - y\|_2^2 + \lambda \|x\|_1 \quad (8)$$

其中, $A_{N(y)} \in \mathbb{R}^{m \times k}$ 是与测试样本 y 邻近的 k 个训练样本组成的矩阵, $\hat{x}_1 \in \mathbb{R}^k$ 为待计算的测试样本的稀疏系数。

3 基于稀疏系数相似性的LSRC方法

k NN算法的优点是思想简单,易于实现,但是当样本维

数比较大时, k NN 的性能会下降。例如, 图 1 为来自 MNIST 数据集中的—个测试样本及其邻近的 5 个训练样本, 使用欧氏距离来衡量样本近邻程度。由于选出的 5 个训练样本中包含数字 8 的有 3 个样本, 因此测试样本数字 5 被误分为数字 8^[15]。

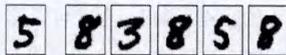


图 1 MNIST 数据集中一个测试样本(最左边)及其 5 个近邻训练样本

稀疏系数反映了样本之间的相关性^[16], 即训练样本表示测试样本的贡献程度, 对于来自同一类别的样本, 它们的稀疏系数应当有较大的相似性, 那么稀疏系数之间的相似性可以用来衡量样本之间的相似性。本文利用稀疏系数的余弦相似性来衡量样本之间的相似性, 以此来选择测试样本的近邻样本。使用 ORL 人脸库中每个人的前 5 幅图像组成训练样本集, 第一个人的第 6 幅图像、第一个人的第 10 幅图像和第二个人的第 6 幅图像作为测试样本, 部分训练样本如图 2 所示, 测试样本及其稀疏系数分别如图 3—图 5 所示。根据余弦相似性计算出图 3 和图 4 稀疏系数的相似性为 0.6529, 而图 3—图 5 稀疏系数的相似性为 -0.0019。这验证了利用稀疏系数相似性来选择近邻样本的可行性。对于图 4 中的样本, 利用 k NN 可以选出同类别的 2 个训练样本, 而利用稀疏系数的相似性可以选出同类别的 3 个训练样本。这在一定程度上可以避免测试样本的错分情况, 进一步提高 LSRC 的识别性能。

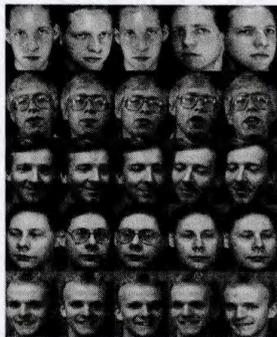


图 2 ORL 人脸库中的部分训练样本

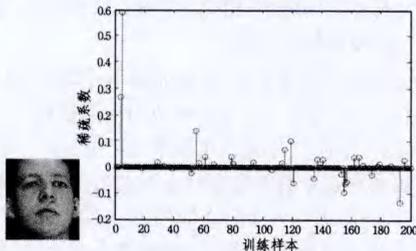


图 3 ORL 人脸库中第一个人的第 6 幅图像及其稀疏系数

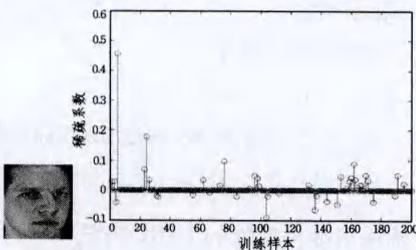


图 4 ORL 人脸库中第一个人的第 10 幅图像及其稀疏系数

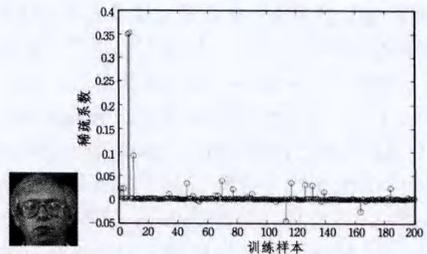


图 5 ORL 人脸库中第二个人的第 6 幅图像及其稀疏系数

综上所述, 本文提出算法的流程如下:

1. 输入: 由训练样本组成的矩阵 A 和测试样本 y , 其中 A 的每一列为一个训练样本;
2. 把矩阵 A 和 y 向量列方向单位化;
3. 分别计算各个训练样本和测试样本 y 在矩阵 A 上的稀疏系数, 依据测试样本的稀疏系数和各个训练样本稀疏系数之间的相似性, 从训练样本中选择 k 个与测试样本邻近的样本, 这些训练样本按列构成矩阵 $A_{N(y)}$;
4. 求解带约束的 l^1 范数最小二乘问题:

$$\hat{x}_1 = \arg \min_x \|A_{N(y)}x - y\|_2^2 + \lambda \|x\|_1$$
5. 计算 y 和重构样本的残差 $r_i(y) = \|y - A\hat{\delta}_i(\hat{x}_1)\|_2$, $i=1, 2, \dots, C$, 其中 C 为样本类别数。
6. 输出: 测试样本的类别: $identity(y) = \arg \min_i r_i(y)$ 。

4 实验验证及分析

本节通过实验来验证所提方法的有效性, 对比方法有 Eigenfaces、Fisherfaces、稀疏表示分类方法(SRC)和局部稀疏表示分类(LSRC)。使用的图像库分别为 ORL 人脸库、Yale 人脸库和 AR 人脸库^[17]。在对图像进行分类前, 使用 PCA 对样本进行降维, 对于 SRC 求解式(6)以及 LSRC 和本文方法求解式(8), 均使用 l_1 - l_2 ^[18] 来求解稀疏系数, 设为 0.005。由于 Fisherfaces 的有效投影维数为 $C-1$ (C 为样本类别数), 因此低于 $C-1$ 的投影维数下 Fisherfaces 的识别率为 N/A 。

4.1 ORL 人脸库

ORL 是基于表情和姿态变化的人脸数据库, 该数据库由 40 个人中每人 10 幅不同姿态和表情的人脸图像组成, 总共 400 张。图像的分辨率为 112×92 , 灰度级为 256。

表 1 不同方法在 ORL 人脸库上的正确识别率

维数	30	40	60	70	80
Eigenfaces	90.42	90	90	90.42	90.83
Fisherfaces	N/A	93.33	93.33	91.25	90.83
SRC	91.25	91.67	93.75	93.33	93.33
LSRC	92.5	92.92	94.58	94.58	95
Proposed	93.33	94.17	94.58	95	95.42

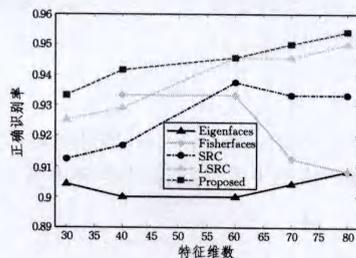


图 6 不同方法在 ORL 人脸库上的识别率曲线图

实验中首先将所有图像缩放为 28×23 , 然后选择每个人的前 4 幅图像组成训练样本, 其余为测试样本。LSRC 和本文方法选择训练样本时 $k=30$, 使用 PCA 对样本进行降维, 不同特征维数下的识别率如表 1 所列, 对应的曲线图如图 6 所示。

4.2 Yale 人脸库

Yale 人脸数据库由 15 个人中每人 11 张不同表情、姿态和光照变化的图像组成, 图像分辨率为 121×160 , 灰度级为 256。实验中首先将所有图像缩放为 46×60 , 然后选择每个人的前 5 幅图像组成训练样本, 其余为测试样本。LSRC 和本文方法选择训练样本时 $k=20$, 使用 PCA 对样本进行降维, 表 2 为不同特征维数下相应算法的识别率, 对应的曲线图如图 7 所示。

表 2 不同方法在 Yale 人脸库上的正确识别率

维数	30	40	50	60	70
Eigenfaces	88.89	88.89	88.89	90	90
Fisherfaces	91.11	91.11	93.33	94.44	94.44
SRC	92.22	94.44	95.56	95.56	96.67
LSRC	93.33	93.33	94.44	94.44	93.33
Proposed	94.44	95.56	95.56	96.67	96.67

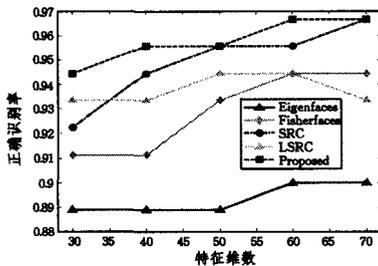


图 7 不同方法在 Yale 人脸库上的识别率曲线图

4.3 AR 人脸库

AR 人脸库有 126 人的超过 4000 幅人脸图像。本次实验使用其中的一个子集(仅考虑光照和表情变化)。这个子集包含 20 个男性和 20 个女性的人脸图像, 实验中首先将所有图像缩放为 50×40 , 然后选择每个人的前 7 幅图像组成训练样本, 其余为测试样本。LSRC 和本文方法选择训练样本时 $k=30$, 使用 PCA 对样本进行降维, 不同特征维数下的识别率如表 3 所列, 对应的曲线图如图 8 所示。

表 3 不同方法在 AR 人脸库上的正确识别率

维数	30	40	50	60	70
Eigenfaces	73.21	73.93	75	74.29	75
Fisherfaces	N/A	97.86	98.57	98.57	98.93
SRC	92.5	93.21	96.43	96.43	97.14
LSRC	94.64	94.64	96.43	96.79	97.17
Proposed	93.57	95	97.14	97.86	97.5

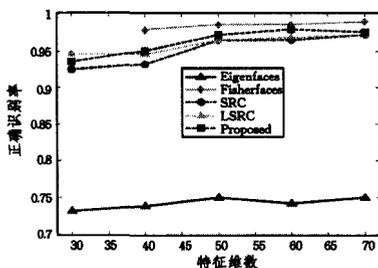


图 8 不同方法在 AR 人脸库上的识别率曲线图

4.4 实验结果及讨论

由以上实验结果可以得到:

1) 通常 SRC 的性能高于 Eigenfaces 和 Fisherfaces, 说明在一定条件下利用稀疏性可以提高人脸识别的准确率; 不过由于 SRC 基于最小重构误差对样本进行分类, 没有有效利用样本的类别信息, 因此在 AR 人脸库上 SRC 的性能不如 Fisherfaces。

2) 在 ORL 和 AR 人脸库上, LSRC 方法优于 SRC 方法, 这说明利用部分训练样本组成字典矩阵可以提高 SRC 的识别性能。而在 Yale 人脸库上, LSRC 的识别效果不是很理想, 这是由于 ORL 人脸库和实验中我们使用的部分 AR 人脸库中的人脸图像的光照、姿态等变化不大, 此时 k NN 可以准确选择同类别样本; 而当样本出现光照、姿态等变化较大时, k NN 的性能会下降, 导致 LSRC 算法的效果不理想。

3) 本文算法总体上优于 SRC 和 LSRC, 这验证了利用稀疏系数之间的相似性衡量样本之间相似性的可行性。由于本文方法未能利用训练样本的类别信息, 因此在 AR 人脸库上本文方法的识别率没有 Fisherfaces 高。

结束语 由于 SRC 使用所有训练样本构成字典矩阵, 当训练样本数目比较多时, 稀疏系数的求解会变得困难。由于测试样本和训练样本稀疏系数之间的相似性可以反映样本之间的相似性, 我们利用稀疏系数之间的相似性来选择部分训练样本。在 ORL 人脸库、Yale 人脸库和 AR 人脸库库上的实验证明了该方法的有效性。不过本文方法属于无监督方法, 如何有效利用训练样本的类别信息以及处理人脸图像遮挡问题是我们今后研究的方向。

参考文献

- [1] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86
- [2] Belhumeur P N, Hespanha J P, Kriegman D. Eigenfaces vs. fisherfaces; Recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720
- [3] He Xiao-fei, Yan Shui-cheng, Hu Yu-xiao, et al. Face recognition using Laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340
- [4] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227
- [5] Wei Chia-po, Chao Yu-wei, Yeh Yi-ren, et al. Locality-sensitive dictionary learning for sparse representation based classification [J]. Pattern Recognition, 2013, 46(5): 1277-1287
- [6] Liu Hui-dong, Yang Ming, Gao Yang, et al. Bilinear discriminative dictionary learning for face recognition[J]. Pattern Recognition, 2014, 47(5): 1835-1845
- [7] Yang Meng, Zhang Lei, Feng Xiang-chu, et al. Fisher discrimination dictionary learning for sparse representation[C]//IEEE International Conference on Computer Vision. 2011: 543-550
- [8] Jiang Zhuo-lin, Lin Zhe, Davis Larry S. Label consistent K-SVD; learning a discriminative dictionary for recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2651-2664

内含子长度的情况下,原始算法和改进算法在 E1 上获得读段定位数量的对比。实验结果表明,改进算法在有效减少执行时间的同时,并不会大幅度降低定位敏感度,即不会大量丢失正确定位位置。在定位搜索空间设置为 400bp、500bp 的情况下,最终所得的读段定位数量并没有明显变动。因此通过融入生物信息可以选择相对小的搜索空间,从而减少算法执行时间,具体情况如表 2 所列。

表 2 不同大小定位搜索空间情况下,改进算法所得到的读段定位数量对比

参考序列名称	搜索空间大小(bp)				
	100	200	300	400	500
AT1G03060.1	1612	1634	1640	1647	1648
AT1G05570.1	2794	3037	3043	3045	3049
AT1G27180.1	433	437	437	437	437
AT1G42470.1	394	449	450	450	450
AT1G48090.1	1710	1769	1800	1823	1829
AT1G50140.1	319	346	349	358	369
AT1G56130.1	154	187	189	189	189
AT1G56140.1	713	791	806	806	807
AT1G80070.1	3791	3916	3933	3933	3934
AT1G80410.1	1301	1399	1416	1436	1452

结束语 本文利用 Hadoop 分布式缓存机制,在读段定位过程中融入生物信息,并且对跨越剪切位的匹配算法进行改进,提出一种改进的 PJuncSeqMap 算法。在拟南芥菜基因数据集上进行验证,结果表明改进的 PJuncSeqMap 算法能够有效减少算法执行时间。未来将进一步研究基于 MapReduce 的基因读段定位算法的性能优化问题,寻找到执行效率更高的解决方案。

参 考 文 献

[1] Jiang H, Wong W H. SeqMap: mapping massive amount of oligonucleotides to the genome[J]. *Bioinformatics*, 2008, 24(20): 2395-2396

[2] Langmead B, Trapnell C, Pop M. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. *Genome Biol*, 2009, 10(3): 25

(上接第 51 页)

[9] Xu Yong, Zhang D, Yang Jian, et al. A two-phase test sample sparse representation method for use with face recognition[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(9): 1255-1262

[10] Zhang Lei, Yang Meng, Feng Xiang-chu. Sparse representation or collaborative representation: Which helps face recognition? [C]// *IEEE International Conference on Computer Vision (ICCV)*. 2011: 471-478

[11] Mi Jian-Xun. Face image recognition via collaborative representation on selected training samples[J]. *Optik-International Journal for Light and Electron Optics*, 2013, 124(18): 3310-3313

[12] Feng Zhi-zhao, Yang Meng, Zhang Lei, et al. Joint discriminative dimensionality reduction and dictionary learning for face recognition[J]. *Pattern Recognition*, 2013, 46(8): 2134-2143

[13] Liu Bao-di, Wang Yu-xiong, Zhang Yu-jin, et al. Learning dictionary on manifolds for image classification[J]. *Pattern Recognition*, 2013, 46(7): 1879-1890

[3] Wang K, Singh D, Zeng Z. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery[J]. *Nucleic Acids Res*, 2010, 38(18): 178

[4] 王曦,汪小我,王立坤,等. 新一代高通量 RNA 测序数据的处理与分析[J]. *生物化学与生物物理进展*, 2010, 37(8): 834-846
Wang Xi, Wang Xiao-wo, Wang Li-kun, et al. A new generation of high-throughput RNA sequencing data processing and analysis[J]. *Progress in Biochemistry and Biophysics*, 2010, 37(8): 834-846

[5] Homer N, Merriman B, Nelson S F. BFAST: an alignment tool for large scale genome resequencing[J]. *PLoS One*, 2009, 4(11): 7767

[6] Smith T F, Waterman M S. Identification of common molecular subsequences[J]. *J Mol Biol*, 1981, 147(1): 195-197

[7] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters[J]. *ACM*, 2008, 51(1): 137-150

[8] 杨晓亮. MapReduce 并行计算应用案例及其执行框架性能优化研究[D]. 南京: 南京大学, 2012
Yang X L. The Application Case Study of MapReduce Parallel Computation and the Optimization of its Runtime Framework [D]. Nanjing: Nanjing University, 2012

[9] Schatz M C. CloudBurst: highly sensitive read mapping with MapReduce[J]. *Bioinformatics*, 2009, 25(11): 1363-1369

[10] 涂金金, 杨明, 郭丽娜. 基于 MapReduce 的基因读段定位算法[J]. *模式识别与人工智能*, 2014, 27(3): 206-212
Tu J J, Yang M, Guo L N. Gene Read Mapping Algorithms Based on MapReduce[J]. *Pattern Recognition and Artificial Intelligence*, 2014, 27(3): 206-212

[11] 刘鹏. 实战 Hadoop: 开启通向云计算的捷径[M]. 北京: 电子工业出版社, 2011
Liu Peng. *Hadoop: open the shortcut to the cloud computing* [M]. Beijing: Electronics Industry Press, 2011

[12] 王立坤. RNA-seq 数据的处理与应用[D]. 吉林: 吉林大学, 2012
Wang Li-kun. *Processing and Application of RNA-seq Data* [D]. Jilin: Jilin University, 2012

[14] Li Chun-guang, Guo Jun, Zhang Hong-gang. Local sparse representation based classification[C]// *International Conference on Pattern Recognition*. 2010: 649-652

[15] Hotta S, Kiyasu S, Miyahara S. Pattern recognition using average patterns of categorical k-nearest neighbors[C]// *Proceedings of the 17th International Conference on Pattern Recognition*. 2004, 4: 412-415

[16] Cheng Bin, Yang Jian-chao, Yan Shui-cheng, et al. Learning with-graph for image analysis[J]. *IEEE Transactions on Image Processing*, 2010, 19(4): 858-866

[17] Martinez A M, Benavente R. The AR face database. CVC Technical Report[R]. 1998, 24

[18] Kim S J, Koh K, Lustig M, et al. A method for large-scale-regularized least squares[J]. *IEEE Journal on Selected Topics in Signal Processing*, 2007, 1(4): 606-617