

结合层次法与主成分分析特征变换的宫颈细胞识别*

赵理莉¹, 孙燎原², 殷建平², 李宽², 印万鹏², 祝恩²

(1. 南通大学 计算机科学与技术学院, 江苏 南通 226091; 2. 国防科技大学 计算机学院, 湖南 长沙 410073)

摘要:对宫颈细胞进行多分类可以自动识别出不同状态的细胞,进而为宫颈癌诊断提供科学依据。在用6种多分类算法进行实验后,选取支持向量机作为基分类器,先用一对一策略训练6个分类器进行3分类,然后再训练1个2分类器,这种二层4分类方法提高了识别准确率。考虑不同层特征模式的差异性,在保证识别性能的同时,每层分类前先采用主成分分析法将原始154维特征变换到低维空间,去除冗余特征,加快识别速度。实验证明,所提层次主成分分析法在宫颈细胞分类中相比6种传统多分类方法有更高的识别准确率,可达90%以上;识别速度也较普通层次法提升了21.31%。

关键词:宫颈涂片图像;特征变换;层次多分类;宫颈细胞识别

中图分类号:TP39 文献标志码:A 文章编号:1001-2486(2017)06-045-06

Cervical cell recognition based on hierarchical method and principal component analysis feature transformation

ZHAO Lili¹, SUN Liaoyuan², YIN Jianping², LI Kuan², YIN Wanpeng², ZHU En²

(1. College of Computer Science and Technology, Nantong University, Nantong 226091, China;

2. College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to recognize multi-class cervical cells automatically, a hierarchical method with PCA (principal component analysis) feature transformation was proposed and this cell recognition could provide the evidence for cervical cancer diagnosis. The cervical cell recognition was treated as a 4-class classification problem. There were two levels in this hierarchical method. First, one-versus-one strategy was used to train 6 SVM (support vector machine) classifiers to do a 3-class classification. Second, abnormal cells in one type of 3 categories were classified by a 2-class SVM. To optimize the feature combination and reduce the running time, a feature transformation method named PCA was adopted to transform the original feature vector into low-dimension feature space. The experiments show that the proposed hierarchical PCA recognition method is faster than the common hierarchical method at a ratio of 21.31%, and can distinguish 4 cervical cell categories better than 6 other traditional methods and achieve above 90% accuracy.

Key words: cervical smear image; feature transformation; hierarchical multi-class classification; cervical cell recognition

在医学图像分析与识别研究领域,宫颈细胞自动识别是一个开放性的热门问题。此类问题的最终目标是找出有异常倾向的和已癌变的细胞个体,这类细胞的数目直接反映出被检测人员的身体状态。因此,定期进行宫颈细胞涂片筛查能够有效降低宫颈癌发病致死概率^[1]。实现宫颈细胞准确自动分析,加快癌细胞识别速度,减轻医务人员读图查找癌细胞的工作量,是该领域学者追求的目标^[2]。

用于医学检测的宫颈细胞图像有以下特点:
①实际应用中,制片成像时,有不可避免的噪声;
②细胞核、细胞质边缘模糊,不易辨别,且细胞重叠粘连比例高;
③异常倾向细胞与癌细胞间的差

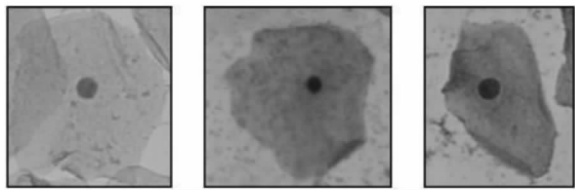
异不显著,准确识别困难;④细胞个体较小,而待检测细胞总体数量大,要求对每个细胞快速分析。鉴于以上特点,宫颈细胞图像分析与普通图像分析有较大区别,进而其在分割、特征提取、对象识别^[3-5]等子问题上也与普通图像分析所用方法不尽相同。

在分割、提取特征后,宫颈细胞图像分析的落脚点在于细胞的识别,而识别本质上是一个多分类问题。本研究中,宫颈细胞分为4类(细胞样例参见图1):正常鳞状上皮细胞、正常柱状细胞、异常倾向细胞和癌细胞。经典分类器,如K近邻(K Nearest Neighbor, KNN)、决策树(Decision Tree, DT)、支持向量机(Support Vector Machine,

* 收稿日期:2016-09-01

基金项目:国家自然科学基金资助项目(61170287,61232016,61303189,61672528)

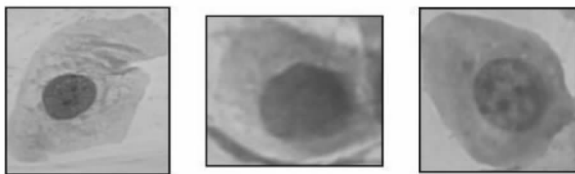
作者简介:赵理莉(1985—),女,河南新乡人,讲师,博士,E-mail:yilinzhaol@nudt.edu.cn



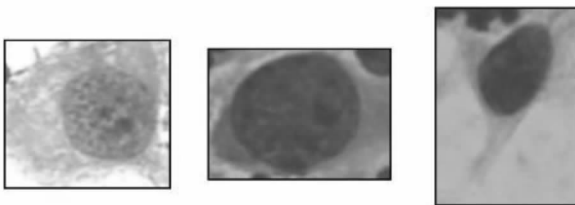
(a) 正常鳞状上皮细胞
(a) Normal squamous cells



(b) 正常柱状细胞
(b) Normal columnar cells



(c) 异常倾向细胞
(c) Abnormal cervical cells



(d) 宫颈癌细胞
(d) Cervical cancer cells

图 1 4 类宫颈细胞样例

Fig. 1 Samples of 4 cervical cell categories

SVM)、随机森林(Random Forest, RF)等都可在一定程度上处理多分类问题。上述分类器中, SVM 的分类原理是在满足最大间隔条件下, 学习出对偶特征空间中正负样本间的最优分割超平面。显然对于多分类问题, 只用一个最优超平面是不能区分多种类型的样本类的, 又因为现实样本分布具有复杂性和多样性, 所以多分类是一个尚未解决的难题。本研究目的是设计一种针对宫颈细胞的准确、快速、较好的识别方法。对现有 6 种经典分类算法实验分析后, 提出一种结合层次法和主成分分析(Principal Component Analysis, PCA)特征变换的识别方法来提高宫颈细胞识别性能。

1 宫颈细胞自动识别

1.1 基于一对一策略的层次多分类

假设有样本数据 $D = \{\mathbf{x}_m, y_m\}, m = 1, \dots, N$, $y_m \in \{1, \dots, K\}$, 其中, \mathbf{x}_m 表示一个样本数据, 样本 \mathbf{x}_m 有 d 维特征, N 为样本数目, y_m 表示样本标签, K 为标签的类别数目。当 $K = 2$ 时, 该分类问题是二分类, 而当 $K > 2$ 时, 该分类问题是多分类。细胞识别是多分类问题, 识别常用方法是训练一组分类器族, 使其能够正确给出测试样本 $\mathbf{x}_t \in \mathbf{R}^d$ 的标签 $y_t \in \{1, \dots, K\}$ 。

目前没有通用的多分类分解求解法^[6], 需要根据问题而设计策略。通常, 多分类求解是通过训练多个基分类器来实现^[7]。其中, 最常用的两种策略为一对一(One-Versus-One, OVO)法和层次方法, 本文结合二者进行细胞识别。

1.1.1 经典支持向量机

简单回顾二分类支持向量机(经典 SVM)。设二分类的类别标签 $y_m \in \{-1, +1\}$, 则 SVM 用来区分样本 \mathbf{x}_m 属于正类或是负类。SVM 算法最终学习到一个满足最大间隔的最优分割超平面。分割超平面能够在核空间中最优化地区分出正负样本类。SVM 的优化目标函数为:

$$\min_{w, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} + c \sum_{m=1}^N \xi_m \quad (1)$$

满足约束:

$$y_m (\mathbf{w}^T \mathbf{x}_m + b) \geq 1 - \xi_m, \xi_m \geq 0, m = 1, \dots, N \quad (2)$$

其中, $\mathbf{w} \in \mathbf{R}^d$ 是用向量形式表示的分割超平面, b 是偏置, ξ_m 是松弛变量, $c > 0$ 是正则化参数。上述优化问题可以转化为一个凸优化问题形式, 求解等价的凸优化问题就可以得到原始问题的解。此外, 为解决非线性不可分问题, 核技巧被引入到优化目标函数式中。核函数用于将特征空间中线性不可分样本数据投影到线性可分空间中, 使得数据可分, 从而能解出分割超平面。

1.1.2 一对一策略

一般来说, 由于二分类的分类决策边界更容易确定, 所以往往将更复杂的多分类转化为二分类问题来求解, 所使用的二分类算法称为基分类器, 本研究使用的基分类器是 SVM。OVO 策略^[8]是将所有类别两两组合成所有可能的二分类组合对, 一个组合对用一个基分类器求解, 最后多分类结果是根据所有基分类器输出对每个类投票, 得分最高的类作为测试样本的输出类别。策略形式化如下:

设有 K 个类别, K 多分类问题的 OVO 策略要求解 $K(K-1)/2$ 个二分类问题。在测试时, 设一个基分类器的输出为 $w_{ij} \in [0, 1]$, 表示对类 i 和类 j , 测试样本属于类 i 的置信度; 而 $w_{ji} = 1 - w_{ij}$, 表示测试样本属于类 j 的置信度。所有基分类器族输出结果可以表示为一个评分矩阵 \mathbf{W} :

$$\mathbf{W} = \begin{pmatrix} - & w_{12} & \cdots & w_{1K} \\ w_{21} & - & & w_{2K} \\ \vdots & & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & - \end{pmatrix} \quad (3)$$

式中, “-”表示置信度最大值。然后根据每个分类器的输出结果投票统计测试样本所属类别, 投票得分最多的类是测试样本的输出类。

1.1.3 层次多分类

层次多分类第一层将所有类分为一级子类, 第二层再将一级子类进一步分为二级子类, 直到能够区分所有类别为止。本工作的层次法分两层, 先将 4 分类转化为基于 OVO 策略的 3 分类问题, 然后对 3 分类结果中的非正常细胞类(包含异常倾向细胞和癌细胞)进一步 2 分类, 最终将所有细胞分为正常鳞状上皮细胞、正常柱状细胞、异常倾向细胞、癌细胞 4 类。图 2 给出了所使用的层次多分类法的示意图。

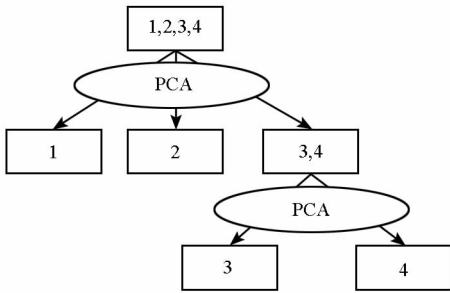


图 2 层次多分类示意图

Fig. 2 Hierarchical multi-class classification

1.2 基于主成分分析的特征变换

为加快识别速度, 在每层分类前对原始特征进行特征变换。特征变换^[9]是降低特征维度的一种思路, 是为了找到一种与原始较高维数据对应的低维流形关系, 使得低维流形能保持原始数据中样本的距离、类别等本质数据信息。PCA 是一种特征变换方法, 其原理是抽取不相关的特征组成新的特征集合来表示原始特征数据。PCA 的形式化定义为:

设原始数据的 N 个样本被表示为 $\mathbf{Z}^T = [Z_1, \dots, Z_N] \in \mathbf{R}^{d \times N}$, 其中, d 为特征维度。先使用奇异值分解 (Singular Value Decomposition, SVD) 求得数

据 \mathbf{Z}^T 的协方差矩阵:

$$\mathbf{\Sigma} = \frac{1}{N-1} \mathbf{Z}^T \mathbf{Z} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T \quad (4)$$

式中: \mathbf{P} 是转换矩阵;

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d), \lambda_1 \geq \dots \geq \lambda_d \geq 0 \quad (5)$$

其中, $\lambda_1, \dots, \lambda_d$ 是协方差矩阵 $\mathbf{\Sigma}$ 的特征值。然后, 前 q 个最大的特征值被选中, 而这些选中成分的累积贡献度可以通过等式(6)计算求得。

$$\xi = \frac{\sum_{i=1}^q \lambda_i}{\sum_{j=1}^d \lambda_j} \quad (6)$$

2 算法细节与描述

2.1 特征归一化

使用 SVM 分类器之前, 对训练和测试数据每维特征进行归一化处理非常重要: ①可避免只考虑大数值特征维在分类中起主导作用却忽略了小数值特征维; ②可避免维度灾难和降低计算代价^[10]。

2.2 识别算法描述

设原始样本数据为 $\mathbf{D} = [D_1, \dots, D_N] \in \mathbf{R}^{d \times N}$, 细胞类别标签为 $1, \dots, K$ 。 $K=1$ 表示正常鳞状上皮细胞类; $K=2$ 表示正常柱状细胞类; $K=3$ 表示异常倾向细胞类; $K=4$ 表示癌细胞类。因此, 基于层次方法和 PCA 变换的宫颈细胞识别算法表述如算法 1 所示。

算法 1 层次 PCA 识别算法

Alg. 1 Hierarchical PCA recognition algorithm

输入: 训练数据 $\mathbf{D}_{\text{train}} = [D_1, \dots, D_{N1}] \in \mathbf{R}^{d \times N1}$,

训练集标签 Y_{train} ,

测试数据 $\mathbf{D}_{\text{test}} = [D_1, \dots, D_{N2}] \in \mathbf{R}^{d \times N2}$

输出: 测试集标签 Y_{test}

1. 第一层 3 分类:
2. 1) 将原始训练集标签处理为 3 类(合并 3 和 4 类)
3. 2) PCA 特征变换
4. 3) 归一化数据
5. 4) 用 OVO 策略, 训练 6 个二分类 SVM 模型
6. 5) 对测试数据进行 3 分类
7. 第二层 2 分类:
8. 1) 抽取原始训练集 3 类和 4 类
9. 2) PCA 特征变换
10. 3) 归一化数据
11. 4) 训练 1 个二分类 SVM 模型
12. 5) 对测试数据第 3 类样本进一步分类
13. 将第一层和第二层测试样本结果合并后输出测试样本的标签 Y_{test}

3 数据、实验与分析

3.1 数据集

实验用公开宫颈细胞数据集 Herlev 来验证提出识别多分类算法的有效性。Herlev 数据集共有 917 张单细胞宫颈细胞图像。其中,正常鳞状上皮细胞 144 张,正常柱状细胞 98 张,异常倾向细胞 525 张,癌细胞 150 张。

3.2 特征集合

特征描述了不同类别细胞的判别特征,宫颈细胞的特征分为三个大类:细胞形态特征、染色质病变特征及区域颜色特征。表 1 列出了每个细胞特征的具体描述。特征抽取需先将细胞图像用 Gap-search MRF 法分割^[11],然后根据分割后图像和原始图像从每个细胞图像中抽取出 154 维特征。

表 1 宫颈细胞特征

Tab.1 Features of cervical cell

(a) 细胞形态特征

(a) Morphology features of each cell

特征维度	描述	特征维度	描述
1	细胞面积	7	核短轴
2	细胞短轴	8	核长轴
3	细胞长轴	9	核偏心率
4	细胞偏心率	10	核圆度
5	细胞圆度	11	核质比
6	核面积	12	核位置

(b) 染色质病变特征

(b) Chromatin pathology features of each cell

特征维度	描述
13 ~ 52	40 维核 Gobar 滤波均值
53 ~ 92	40 维核 Gobar 滤波方差
93 ~ 118	26 维旋转不变 LBP ^[12]

(c) 区域颜色特征

(c) Region intensity features of each cell

特征维度	描述	特征维度	描述
119 ~ 121	核 RGB 均值	137 ~ 139	质 RGB 均值
122 ~ 124	核 RGB 方差	140 ~ 142	质 RGB 方差
125 ~ 127	核 HSV 均值	143 ~ 145	质 HSV 均值
128 ~ 130	核 HSV 方差	146 ~ 148	质 HSV 方差
131 ~ 133	核 LAB 均值	149 ~ 151	质 LAB 均值
134 ~ 136	核 LAB 方差	152 ~ 154	质 LAB 方差

3.3 主成分数目选取

层次识别方法各个层次主成分选取的数目是

根据实验得到。图 3 给出了主成分数目选取依据。根据 50 次重复实验结果,选用 80 作为特征变换后的主成分数目。

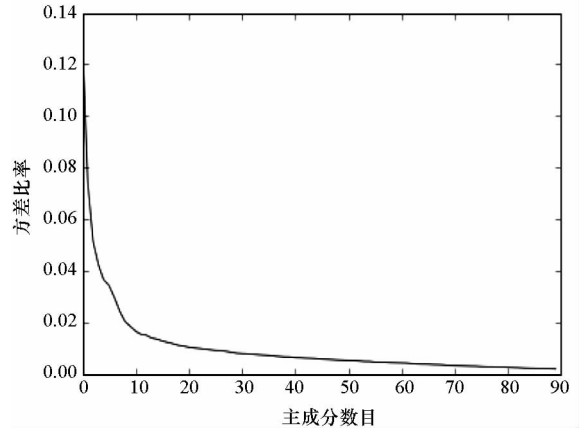


图 3 前 90 个主成分的方差比率

Fig.3 Variance ratio of top 90 principal components

3.4 分类结果

采用的识别评估指标有:准确率、精度、召回率、F-score、混淆矩阵。

主要进行了 2 组实验:①测试分类准确率的实验。该实验是对所提方法与常用多分类器的准确率进行比较。在使用 Gap-search MRF 对细胞分割后,进一步用层次法、层次 PCA 方法和 6 种分类器对细胞进行多分类。6 种分类器为逻辑斯谛回归 (Logistic Regression, LR)、K 近邻分类 (KNN)、随机森林 (RF)、决策树 (DT)、支持向量机 (SVM) 以及梯度下降决策树 (Gradient Boosting Decision Tree, GBDT)。②测试分类时间的实验。该实验是各种分类方法速度的比较。

为了避免过拟合,实验训练集和测试集使用 9 : 1 的划分比例,测试集样本是用伪随机数选取得到的。

表 2 的实验是基于细胞的 154 维特征,用 6 种多分类器对细胞直接进行 4 分类。这 6 种方法中 GBDT 分类准确率最高,SVM 略低,二者准确率基本相当。

表 2 6 种分类器 4 分类结果

Tab.2 4-class classification results of 6 classifiers

分类器	测试准确率/%	精度	召回率	F-score
LR	78.26	0.80	0.78	0.78
KNN	76.09	0.77	0.76	0.76
RF	82.61	0.83	0.83	0.83
DT	68.48	0.71	0.68	0.69
SVM	85.70	0.86	0.86	0.86
GBDT	86.96	0.87	0.87	0.87

为进一步分析 GBDT 和 SVM,分别给出两种分类器的混淆矩阵(见表 3 和表 4)。混淆矩阵的对角线元素表示正确分类的样本,而非对角线元素为错误分类的样本。

表 3 GBDT 4 分类混淆矩阵

Tab.3 Confusion matrix of GBDT for 4-class classification

	类 1	类 2	类 3	类 4	#测试细胞
类 1	13	0	0	0	13
类 2	0	6	1	1	8
类 3	0	1	38	3	42
类 4	0	2	4	23	29

注:类 1,2,3,4 分别为正常鳞状上皮细胞、正常柱状细胞、异常倾向细胞、癌细胞。

表 4 SVM 4 分类混淆矩阵

Tab.4 Confusion matrix of SVM for 4-class classification

	类 1	类 2	类 3	类 4	#测试细胞
类 1	10	0	3	0	13
类 2	0	4	4	0	8
类 3	0	0	60	0	60
类 4	0	0	6	5	11

由表 3 和表 4 对比知,SVM 对于第 3 类细胞样本的区分度优于 GBDT,而对于第 2 类和第 4 类样本的区分度略弱于 GBDT。

为选取较优的第一层 3 分类分类器,继续用 6 种分类器进行 3 分类,实验结果见表 5。由表 5 知,SVM 是准确率最高的分类方法;而 GBDT、DT、RF 训练准确率虽高,但是测试准确率低,存在过拟合问题。因此,选取 SVM 作为第一层 3 分类的基分类器。

表 5 第一层 3 分类结果

Tab.5 3-class classification results of first level

分类器	训练准确率/%	测试准确率/%	精度	召回率	F-score
LR	93.21	93.48	0.93	0.93	0.93
KNN	95.27	90.22	0.90	0.90	0.90
RF	99.64	91.30	0.91	0.91	0.90
DT	100.00	89.13	0.90	0.89	0.90
SVM	99.00	98.57	0.98	0.98	0.98
GBDT	100.00	91.30	0.91	0.91	0.90

为将非正常细胞进一步分为异常倾向类和癌细胞类,用 6 种分类器对非正常细胞再进行 2 分

类,结果见表 6。表 6 中 SVM 的分类测试准确率最高,因此第二层基分类器也选用 SVM。

表 6 第二层 2 分类结果

Tab.6 2-class classification results of second level

分类器	训练准确率/%	测试准确率/%	精度	召回率	F-score
LR	94.07	82.22	0.85	0.85	0.85
KNN	87.78	77.78	0.83	0.83	0.83
RF	98.89	85.19	0.84	0.84	0.84
DT	100.00	77.04	0.79	0.79	0.79
SVM	99.04	90.19	0.91	0.91	0.91
GBDT	100.00	83.70	0.88	0.88	0.88

表 7 为 3 种分割方法与所提 2 种分类方法组合的识别结果。由表 7 知:①结合表 2 可得出分类准确率的实验效果——所提出的层次方法和层次 PCA 方法比其他 6 种分类方法(表 2 中)的 4 分类准确率都高;②层次法比层次 PCA 方法分类的准确率略高;③基于阈值分割和分水岭分割的分类方法比基于 Gap-search MRF 分割的分类方法准确率更低,可知不同分割方法对细胞分类准确率有较大影响。

表 7 所提 2 种分类法在 3 种分割方法下的 4 分类结果

Tab.7 4-class classification results of 2-proposed recognition methods with 3 segmentation approaches

分割法 + 分类器	测试准确率/%	精度	召回	F-score
Gap-search MRF + 层次法	91.86	0.91	0.91	0.91
Gap-search MRF + 层次 PCA	90.13	0.90	0.91	0.90
阈值 + 层次法	67.91	0.67	0.68	0.67
阈值 + 层次 PCA	65.13	0.63	0.62	0.62
分水岭 + 层次法	73.69	0.72	0.71	0.71
分水岭 + 层次 PCA	71.52	0.71	0.70	0.71

测试识别时间实验结果见表 8,表中共有 7 种分类方法的运行时间结果,其中运行时间为训练与测试所用时间之和。在 7 种方法中,分类速度最慢的是 GBDT,最快的是 KNN 和 DT。在医学检测实际中,细胞识别首先要保证正确率,其次尽可能提高速度。根据上述原则,层次法准确率最高,但是该方法的识别速度较慢(0.231 40 s),为加快速度,提出了层次法和 PCA 结合的方法。层

次 PCA 方法的分类时间为 0.182 10 s, 比 GBDT、层次法和 RF 快。层次 PCA 方法虽比 KNN、DT 和 LR 三种方法识别速度慢, 但层次 PCA 法的识别准确率更高。

综上, 通过分类准确率和速度的实验知, 所提出的层次法和层次 PCA 方法比其他 6 种方法的分类准确率高, 两者在 Gap-search MRF 分割后的准确率都达 90% 以上; 所提出的层次 PCA 方法比层次法在分类准确率方面仅下降了 1.73%, 但在分类速度方面却提升了约 21.31%。因此, 层次 PCA 方法是综合考虑分类准确率和识别时间的较优方案。

表 8 7 种识别方法的运行时间

Tab. 8 Running time of 7 recognition methods

分类器	运行时间/s
LR	0.090 00
KNN	0.080 00
RF	0.190 00
DT	0.080 00
GBDT	1.310 00
层次法	0.231 40
层次 PCA 法	0.182 10

4 结论和讨论

对宫颈涂片图像中细胞的正确识别是判断病患癌症程度的重要依据, 也是一种研究癌细胞病变机理的重要手段。为了更准确、快速识别不同类型的宫颈细胞, 提出一种基于层次方法和 PCA 特征变换的识别方法, 实验证明, 该方法是综合考虑准确率和识别时间的较优方案。

本工作使用的 PCA 是一种线性特征转换方法, 而有学者提出核 PCA^[13] 可使用非线性变换的核函数将特征投影到高维对偶空间, 最后在对偶空间求解出目标函数的优化解。因此, 总结进一步需改进的方面有: ①使用核学习方法, 生成适用的样本核矩阵, 将特征变换到高维易分空间, 在新空间对细胞进行更准确的分类, 以提升分类准确率。其中, 核函数(比如多项式核、高斯核等)的选取及与核相关的参数的选择是求解的关键。②核学习算法中矩阵运算的计算量大, 将核学习算法并行化, 充分发挥计算机 CPU 和 GPU 的硬件运算能力^[14-15], 以加快识别速度。

参考文献 (References)

[1] World Health Organization. Comprehensive cervical cancer

control: a guide to essential practice second edition [R]. WHO Press, 2014.

- [2] Chen S, Zhao M, Wu G, et al. Recent advances in morphological cell image analysis [J]. *Computational and Mathematical Methods in Medicine*, 2012(1): 143 - 156.
- [3] 翟永平, 周东翔, 刘云辉. 基于颜色及梯度统计特征的结核杆菌目标识别 [J]. *国防科技大学学报*, 2012, 34(5): 146 - 152.
ZHAI Yongping, ZHOU Dongxiang, LIU Yunhui. Recognition of mycobacterium tuberculosis in microscopic images based on color and gradient feature [J]. *Journal of National University of Defense Technology*, 2012, 34(5): 146 - 152. (in Chinese)
- [4] 史殿习, 李勇谋, 丁博. 无监督特征学习的人体活动识别 [J]. *国防科技大学学报*, 2015, 37(5): 128 - 134.
SHI Dianxi, LI Yongmou, DING Bo. Unsupervised feature learning for human activity recognition [J]. *Journal of National University of Defense Technology*, 2015, 37(5): 128 - 134. (in Chinese)
- [5] Marinakis Y, Dounias G, Jantzen J. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification [J]. *Computers in Biology and Medicine*, 2009, 39(1): 69 - 78.
- [6] Lorena A C, de Carvalho A C P L F, Gama J M P. A review on the combination of binary classifiers in multiclass problems [J]. *Artificial Intelligence Review*, 2008, 30(1): 19 - 37.
- [7] Fürnkranz J. Round robin classification [J]. *Journal of Machine Learning Research*, 2002, 2(4): 721 - 747.
- [8] Knerr S, Personnaz L, Dreyfus G. Single-layer learning revisited: a stepwise procedure for building and training a neural network [C]//*Proceedings of the NATO Conference on Neurocomputing*, 1990: 41 - 50.
- [9] Kusiak A. Feature transformation methods in data mining [J]. *IEEE Transactions on Electronics Packaging Manufacturing*, 2001, 24(3): 214 - 221.
- [10] Wang G, Yin S. Quality-related fault detection approach based on orthogonal signal correction and modified PLS [J]. *IEEE Transactions on Industrial Informatics*, 2015, 11(2): 398 - 405.
- [11] Zhao L, Li K, Wang M, et al. Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF [J]. *Computers in Biology and Medicine*, 2016, 71: 46 - 56.
- [12] Ojala T, Pietikäinen M, Mäenpää T, et al. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971 - 987.
- [13] Zhang C, Nie F, Xiang S. A general kernelization framework for learning algorithms based on kernel PCA [J]. *Neurocomputing*, 2010, 73(4/5/6): 959 - 967.
- [14] Liao X, Xiao L, Yang C, et al. MilkyWay-2 supercomputer: system and application [J]. *Frontiers of Computer Science Selected Publications*, 2014, 8(3): 345 - 356.
- [15] Guo S, Dou Y, Lei Y, et al. An efficient multi-standard QC-LDPC decoder based on the row-layered decoding algorithm [J]. *IEICE Electronics Express*, 2015, 12(13): 20150356 - 20150361.