

# 结合信息能度量的核形态差异学习\*

刘松华<sup>1,2</sup>, 丁彩英<sup>1,3,4</sup>, 刘建生<sup>1</sup>

(1. 江西理工大学理学院, 江西 赣州 341000; 2. 西安电子科技大学计算机学院, 西安 710071; 3. 兰州大学交叉学科研究中心, 兰州 730000; 4. 中国科学院物理研究所, 北京 100871)

**摘 要:** 利用核映射及非局部均值降噪特性构造相似性度量, 即信息能度量。通过对细胞表型图进行特征映射并结合信息能度量得到图像特征的形态差异目标函数, 依据梯度上升优化获取最优度量矩阵, 建立基于核方法的形态差异学习模型。该模型特点在于: 不仅考虑了各类样本的相似性, 同时在降低噪声影响形态差异学习的过程中充分利用了图像的高阶统计量和非线性特征。实验结果表明, 该核化算法灵敏度更高, 且具有较好的鲁棒性, 能有效应用于临床诊断。

**关键词:** 图像分类; 核方法; 降噪; 形态差异

**中图分类号:** TP391.41

**文献标志码:** A

**文章编号:** 1001-3695(2012)04-1341-04

doi:10.3969/j.issn.1001-3695.2012.04.039

## Kernel morphological differences algorithm based on informative energy metric

LIU Song-hua<sup>1,2</sup>, DING Cai-ying<sup>1,3,4</sup>, LIU Jian-sheng<sup>1</sup>

(1. College of Science, Jiangxi University of Science & Technology, Ganzhou Jiangxi 341000, China; 2. School of Computer Science & Technology, Xidian University, Xi'an 710071, China; 3. Center of Interdisciplinary, Lanzhou University, Lanzhou 730000, China; 4. Institute of Physics Chinese Academy of Sciences, Beijing 100871, China)

**Abstract:** Firstly, this paper defined the similarity measure by kernel mapping and non-local means, named as informative energy metric (IEM). Secondly, combined the feature mapping of cell images with the IEM and an objective function for morphological differences was computed. Thirdly, it obtained the optimal metric matrix by gradient ascent algorithm and constructed a morphological difference learning model based on kernel methods. The characteristics of this model were that it could measure the similarity between each sample pairs and mine the high order statistics and nonlinear features. Experimental results show that this method is more sensitive and robust, it can be used in disease diagnosis.

**Key words:** image classification; kernel methods; denoising; morphological differences

## 0 引言

临床实验中, 细胞表型图的特征形态变化有助于理解药物在生物体中所起的作用, 是重要的生物学指标<sup>[1]</sup>。但细胞表型图在采集、分析的过程中会受随机抽样和噪声的影响, 因此细胞表型图形态差异便成为图像分类<sup>[2]</sup>和数据分析中的一个典型问题。

近年来, 细胞形态差异学习已经成为药物研究的重要工具<sup>[3,4]</sup>。文献[5]利用支持向量机在核空间中建立超平面, 寻找细胞群体的最大分离方向, 将未知作用的新化合物与已知作用的化合物关联, 进而推断新化合物的药物作用。文献[6]将 AdaBoost 算法与神经网络结合设计了分类性能较好的系综分类器, 能有效进行细胞表型图的分类任务和形态差异分析。然而目前的分析方法仅在细胞级别来区分形态差异, 对于亚细胞级别的部分变化或者混合样本仍然没有可靠的方法。针对这一情况, 文献[7]给出一种基于谱图理论的定量标度细胞表型图形态差异的方法, 采用欧氏距离学习样本之间的相似性, 定

义图转移能量 (graph transition energy) 来定量描述细胞表型图之间的形态差异, 同时加入 Cheeger 常量<sup>[8]</sup>抑制噪声。然而该方法不便于挖掘细胞表型图中特征的高阶统计量和非线性特征, 且样本在高维空间中不一定符合欧氏度量, 同时 Cheeger 常量在降噪的同时没有考虑特征的全局相似性。因此需要设计一种合适的度量, 兼容降噪特性和相似性度量, 避免基于欧氏距离的度量在高维数据中的局限。

本文结合非局部均值 (non-local means, NL means) 降噪运算<sup>[9,10]</sup>特性, 借助核方法设计并学习合适的相似性度量, 使其不仅能充分表示高维数据的距离信息, 挖掘数据中的高阶统计量和非线性特征, 同时能有效提高算法抵抗噪声<sup>[11,12]</sup>的能力。

## 1 核形态差异学习模型

核方法能有效解决数据的线性不可分问题, 且其复杂度不依赖于数据的维数, 因此能较好地应用于医学图像数据。本文在核空间中考虑细胞表型图的特征样本, 并学习最优度量矩阵, 能在一定程度上提高形态差异学习算法的性能。

**收稿日期:** 2011-10-20; **修回日期:** 2011-11-26 **基金项目:** 江西理工大学博士启动基金资助项目; 江西省教育厅青年自然科学基金资助项目; 国家自然科学基金资助项目 (61070137)

**作者简介:** 刘松华 (1981-), 男, 山西河曲人, 博士, 主要研究方向为智能信息处理、生物信息处理 (sooh.liu@gmail.com); 丁彩英 (1981-), 女, 博士, 主要研究方向为非线性动力学、生物信息处理; 刘建生 (1959-), 男, 副教授, 主要研究方向为图像处理、机器学习等。

### 1.1 信息能量度

设  $x_{ci}$  是输入空间  $X$  中第  $c$  类第  $i$  个样本, 当计算与类别无关时, 忽略类别简记为  $x_i$ 。  $C$  是样本对应的类别集合,  $N_c$  表示样本类别总数,  $J_c$  表示第  $c$  类中样本数目, 则  $c \in [1, N_c]$ ,  $i \in [1, J_c]$ 。  $y_{ci}$  为  $x_{ci}$  经过核映射  $\varphi: X \rightarrow Y$  后的核空间  $Y$  中样本, 且  $y_{ci} = \varphi(x_{ci})$ 。 简单起见, 采用与输入空间相同记法, 考虑其  $k$  个最近邻,  $N_c$  表示最近邻中样本类别数目,  $J_c$  表示最近邻中同类样本数目。

文献[7]中采用欧氏距离度量样本相似性, 并给出了一种图转移能量函数, 根据本文写法记为

$$E(\sigma) = \frac{1}{N} \sum_{i,j} T_{ij} I'(x_i \neq x_j) \quad (1)$$

其中:  $N$  为样本总数;  $\sigma$  为待定参数;  $T_{ij} = \frac{w_{ij}}{\sum_l w_{il}}$  且  $w_{ij} = \exp(-\frac{d}{\sum_{d=1}^D \frac{(x_i - x_j)^2}{\sigma_d^2}})$ ;  $I'(x_i \neq x_j)$  为指示函数,  $x_i$  与  $x_j$  的类别不同时函数值为 1, 反之为 0;  $D$  为样本维数。 通过参数  $\sigma$  的学习, 当能量  $E(\sigma)$  最小时, 表示不同类的样本距离越远。

本文在核空间内考虑样本之间的相似性, 进而学习合适的核相似能量函数, 记为  $E$ 。

对于任一样本  $y_{ci}$ , 为了兼容 NL means, 考虑其  $k$  个最近邻, 近邻中与  $y_{ci}$  类别不同的样本其核相似能量记为  $E_{p \neq c}(y_{ci})$ , 与式(1)相同, 不同类别样本远离时, 能量值较低。 由于式(1)只统计不同类样本对能量目标函数的影响, 而忽略了同类样本之间的互相影响以及可能存在的噪声问题。 因此本文将式(1)进行改进, 考虑其同类样本对能量函数的影响, 设近邻中与  $y_{ci}$  类别相同的样本其核相似能量记为  $E_c(y_{ci})$ 。 为了抵抗噪声的影响, 同类样本应该尽量靠近, 即能量值越高性能越好。 因此本文提出信息能量度, 分别计算为

$$E_{p \neq c}(y_{ci}) = \sum_{p=1}^{N_c} \sum_{l=1}^{J_p} G(y_{pl} - y_{ci}, 2\sigma^2 I) \quad (2)$$

$$E_c(y_{ci}) = \sum_{j=1}^{J_c} G(y_{cj} - y_{ci}, 2\sigma^2 I) \quad (3)$$

其中:  $G(y, \sigma^2) = \exp(-\frac{1}{2\sigma^2} y^T y)$  为高斯核函数;  $I$  为单位矩阵;  $\sigma$  为待定参数。 由于核映射  $\varphi$  没有显式表示, 因此  $y_{pl} - y_{ci}$  通过核矩阵  $K$  表示为

$$y_{pl} - y_{ci} = A(\sum_{s,t} (K(x_s, x_{ci}) - K(x_t, x_{ci}))) \quad (4)$$

其中:  $A$  为系数矩阵,  $K(\cdot, \cdot)$  为核矩阵元素。

与文献[7]中式(1)比较, 本文信息能具有如下特点: a) 同时考虑同类和不同类样本之间的影响; b) 充分利用核方法特性, 在核空间中考虑样本之间的距离度量, 避免基于欧氏距离的度量在输入空间的局限。

### 1.2 形态差异学习模型

文献[7]中的形态差异学习模型针对参数  $\sigma$  进行调整, 而  $\sigma$  的选取在核模型中仍然是难点。 因此本文将其转换为度量矩阵  $A$  的学习,  $\sigma$  采用文献[13]中的方法确定。

对于任一样本  $y_{ci}$ , 由于其受相同类别和不同类别的共同影响, 因此根据式(2)和(3)得出形态差异学习模型为

$$E(y_{ci}) = \alpha E_c(y_{ci}) - (1 - \alpha) E_{p \neq c}(y_{ci}) \quad (5)$$

其中:  $\alpha = [(1 - \frac{J_c}{k+1})^2 + \sum_{p=1, p \neq c}^{N_c} (\frac{J_p}{k+1})^2]$ ,  $\alpha$  为相同类别和不同类

别样本对  $y_{ci}$  的影响参数。 第一项表示对于  $y_{ci}$ , 其他所有样本先验概率的共同影响; 第二项表示其他各类样本各自先验概率的影响之和。 根据式(4), 对于所有样本式(5)形态差异学习模型可以定义为  $A$  的函数

$$E(A) = \frac{1}{N^2} \sum_{c=1}^{N_c} \sum_{i=1}^{J_c} (\alpha E_c(y_{ci}) - (1 - \alpha) E_{p \neq c}(y_{ci})) \quad (6)$$

最大化式(6)可以得到特征空间转换的度量矩阵  $A$ , 本文采用梯度上升算法求解, 则目标函数式(6)的梯度为

$$\frac{\partial E}{\partial A} = \frac{1}{N^2} \sum_{c=1}^{N_c} \sum_{i=1}^{J_c} (\alpha \frac{\partial E_c}{\partial A} - (1 - \alpha) \frac{\partial E_{p \neq c}}{\partial A}) \quad (7)$$

其中:  $\frac{\partial E_c}{\partial A} = \frac{\partial E_c}{\partial y_{ci}} \frac{\partial y_{ci}}{\partial A}$  与  $\frac{\partial E_{p \neq c}}{\partial A} = \frac{\partial E_{p \neq c}}{\partial y_{ci}} \frac{\partial y_{ci}}{\partial A}$  可由链式规则分别得出, 且梯度算子根据式(2)和(3)计算为

$$\frac{\partial E_c}{\partial y_{ci}} = \sum_{j=1}^{J_c} G(y_{cj} - y_{ci}, 2\sigma^2 I) \frac{(y_{cj} - y_{ci})}{\sigma^2} \quad (8)$$

$$\frac{\partial E_{p \neq c}}{\partial y_{ci}} = \sum_{p=1}^{N_c} \sum_{l=1}^{J_p} G(y_{pl} - y_{ci}, 2\sigma^2 I) \frac{(y_{pl} - y_{ci})}{\sigma^2} \quad (9)$$

$$\frac{\partial y_{ci}}{\partial A} = \sum_s K(x_s, x_{ci}) \quad (10)$$

利用式(4)将式(8~10)代入式(7)得出目标函数的梯度, 根据梯度上升算法得出最优的度量矩阵  $A$ , 代入式(6)则可确定形态差异学习模型的能量值。

本文形态差异学习模型式(6)与文献[7]的主要区别有:

- 对于能量函数, 低能量表示不同类别的样本远离, 为了应用于分类, 本文同时考虑同类样本聚集;
- 通过核方法将核参数学习问题转换为度量矩阵的学习, 便于对测试样本进行计算;
- 本文模型在求解度量矩阵的过程中加入了 NL means 降噪模型, 便于考虑样本的全局相似性, 因此具有较好的降噪性能。

### 1.3 模型鲁棒性能分析

在形态差异学习模型中文献[7]采用 Cheeger 常量能有效保留分类需要的边缘特征, 然而由于细胞表型图中特征具有全局相似性, 因此需要进一步保留有益于分类的全局特征。 分析本文提出的形态学习模型可知, 本文信息能量度[14]模型具有较好的鲁棒性能。

**定理 1** 给定核空间  $Y$  上的形态差异学习模型  $E$ , NL means 可由  $E$  推出, 且  $E$  为 NL means 的核扩展形式, 因此  $E$  具有较好的鲁棒性。

**证明** 文献[9]中提出的 NL means 降噪模型根据本文写法记为

$$NL(i) = \sum_j u(i, j) x_j \quad (11)$$

其中:  $u(i, j) = \frac{1}{Z(i)} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$ ;  $NL(i)$  为降噪后的样本值, 且  $Z(i) = \sum_j \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$ 。

根据本文形态差异学习模型中式(8)(9)可知,  $G(y_{cj} - y_{ci}, 2\sigma^2 I) \frac{(y_{cj} - y_{ci})}{\sigma^2}$  等价于上述 NL means 降噪模型的核扩展形式。

区别在于前者采用  $Z(i)$  作为规范因子, 而本文模型采用  $\frac{1}{N^2}$  作为规范因子。 由上可知, 本文的模型  $E$  兼容了 NL means 良好的降噪性能, 且为 NL means 降噪模型的核扩展形式。

**定理2** 设式(5)中参数 $\alpha$ 为常量,则当 $k = N - 1$ ,目标函数式(5)等价于互信息度量(mutual information, MI)。

**证明** 文献[17]中的互信息 MI 采用 Renyi 熵计算,根据本文定义,将文献[17]中 MI 记为

$$MI = V_{IN} + V_{ALL} - 2V_{BTW} \quad (12)$$

式(12)中的量分别计算为

$$V_{IN} = \frac{1}{N^2} \sum_{p=1}^{N_c} \sum_{l=1}^{J_p} \sum_{k=1}^{J_p} G(y_{pk} - y_{pl}, 2\sigma^2 I) \quad (13)$$

$$V_{ALL} = \frac{1}{N^2} \left( \sum_{p=1}^{N_c} \left( \frac{J_p}{N} \right)^2 \right) \sum_{k=1}^N \sum_{l=1}^N G(y_k - y_l, 2\sigma^2 I) \quad (14)$$

$$V_{BTW} = \frac{1}{N^2} \sum_{p=1}^{N_c} \sum_{j=1}^{J_p} \sum_{k=1}^N G(y_{pj} - y_k, 2\sigma^2 I) \quad (15)$$

将式(13)~(15)代入式(12)可得互信息 MI 的值。

考虑本文目标函数式(5),记为如下形式:

$$E(y) = \frac{1}{N} \sum_{c=1}^{N_c} \sum_{i=1}^{J_c} \alpha E_c(y_{ci}) - (1 - \alpha) E_{p \neq c}(y_{ci})$$

考虑核空间内与样本 $y_{ci}$ 的 $N - 1$ 个其他样本,则可以将 $\alpha$ 常量中 $k$ 赋为 $N - 1$ 。参数 $\alpha$ 为

$$\alpha = 1 + \left( \frac{J_c}{N} \right)^2 + \sum_{p=1}^{N_c} \left( \frac{J_p}{N} \right)^2 - \frac{2J_c}{N} \quad (16)$$

将参数代入 $E(y)$ 得

$$E(y) = \frac{1}{N^2} (E^1 + E^2 + E^3)$$

其中: $E^1$ 、 $E^2$ 分别计算如下,且记 $G(\cdot - \cdot, 2\sigma^2 I)$  as  $G(\cdot - \cdot)$ :

$$\begin{aligned} E^1 &= \sum_{c=1}^{N_c} \sum_{i=1}^{J_c} \sum_{j=1}^{J_c} G(y_{cj} - y_{ci}) \\ E^2 &= \sum_{c=1}^{N_c} \left( \frac{J_c}{N} \right)^2 \sum_{i=1}^{J_c} \left( \sum_{j=1}^{J_c} G(y_{cj} - y_{ci}) + \sum_{p=1, p \neq c}^{N_c} \sum_{l=1}^{J_p} G(y_{pl} - y_{ci}) \right) + \\ &\quad \sum_{p=1, p \neq c}^{N_c} \left( \frac{J_p}{N} \right)^2 \sum_{c=1}^{N_c} \left( \sum_{i=1}^{J_c} G(y_{cj} - y_{ci}) + \sum_{l=1}^{J_p} G(y_{pl} - y_{ci}) \right) \\ E^3 &= -2 \sum_{c=1}^{N_c} \sum_{i=1}^{J_c} \sum_{j=1}^{J_c} G(y_{cj} - y_{ci}) + \sum_{p=1, p \neq c}^{N_c} \sum_{l=1}^{J_p} G(y_{pl} - y_{ci}) \end{aligned}$$

对于 $E^1$ ,计算同类信息能,等价于式(12)中的 $V_{IN}$ ,除了规范因子 $\frac{1}{N^2}$ 的区别,则 $E^1 = \frac{1}{N^2} V_{IN}$ 。

对于 $E^2$ ,项 $\sum_{j=1}^{J_c} G(y_{cj} - y_{ci})$ 计算类 $c(c \neq p)$ 的信息能,另外一项 $\sum_{l=1}^{J_p} G(y_{pl} - y_{ci})$ 计算类 $p(p \neq c)$ 的信息能,则 $E^2$ 修改为

$$E^2 = \left( \sum_{p=1}^{N_c} \left( \frac{J_p}{N} \right)^2 \right) \sum_{k=1}^N \sum_{l=1}^N G(y_k - y_l, 2\sigma^2 I)$$

对于 $E^3$ ,其中的项计算类 $c(c \neq p)$ 和类 $p(p \neq c)$ 的信息能,则合并为 $\sum_{i=1}^{J_c} \sum_{k=1}^N G(y_k - y_{ci}, 2\sigma^2 I)$ ,因此 $E^3$ 计算为

$$E^3 = -2 \sum_{p=1}^{N_c} \sum_{j=1}^{J_p} \sum_{k=1}^N G(y_{pj} - y_k, 2\sigma^2 I)$$

将 $E^1$ 、 $E^2$ 和 $E^3$ 代入 $E(y) = \frac{1}{N^2} (E^1 + E^2 + E^3)$ ,由此得

$$E(y) = V_{IN} + V_{ALL} - 2V_{BTW}$$

综上,当 $k = N - 1$ ,本文基于信息能的目标函数等价于互信息度量。

由定理2可知,本文采用信息能度量构造形态差异学习目标函数,能有效挖掘图像特征中的高阶统计量与非线性特征,并能推出互信息度量,具有较好的理论基础。

## 2 实验结果与分析

为测试提出的细胞表型图形态差异学习模型的有效性,本文采用2D Hela细胞表型图数据<sup>[15]</sup>,共862幅10类细胞表型图。实验分为定量分析和分类测试两部分。为了方便比较,文献[7]的模型记为QEM(quantification energy model),本文模型记为Kernel-SEM(kernel-similarity energy model)。

### 2.1 形态差异定量分析比较

本文采用文献[7]的数据,利用梯度上升优化式(7),根据获取的最优度量矩阵 $A$ ,用式(6)定量描述细胞表型图的形态差异。实验取100次测试结果近似能量函数分布,结果如图1所示。图1中的数据包括MI(mostly impact)和MC(mostly fragmented completely)<sup>[7]</sup>,分别表示细胞表型图中线粒体分裂的最初图像和最后图像。图1中0% MC表示仅用MI数据,50% MC表示将MI与MC等比例混合,100% MC表示仅用MC数据。其中,图1(a)为QEM运行结果,图1(b)为Kernel-SEM所得结果。

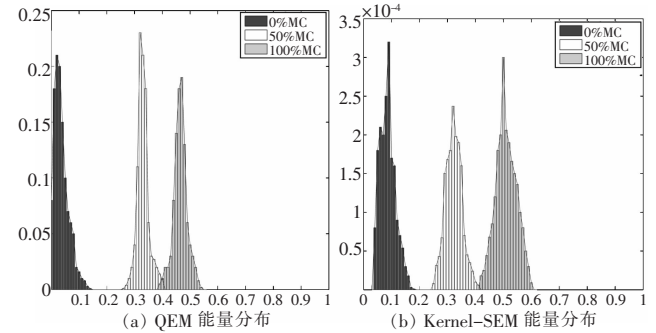


图1 能量模型定量分析性能比较

由图1可知,QEM的100% MC与50% MC的能量分布有较大的重叠,因此会导致部分MC样本与混合样本无法区分,而图1(b)则重叠较小。同时图1(a)的能量值较大,而图1(b)采用Kernel-SEM后,能量的数量级为 $10^{-4}$ 。因此Kernel-SEM能较好地地区分样本间细微的差距,用于定量分析较为灵敏,有利于区分样本混合程度,进一步用于分类。综上可知,本文的Kernel-SEM具有较好的灵敏性,能有效区别细胞表型图特征之间的差异,在临床实验中具有较高的应用价值。

### 2.2 鲁棒性能分析比较

本节实验分为两部分,采用SLF16<sup>[7]</sup>细胞表型图数据。为了避免过拟合现象,采用10折交叉验证抽取样本,抽取60%作为训练样本,40%作为测试样本,同时将测试样本按比例分为有类别和无类别两种。为了与文献[7]比较,采用相同的半监督分类器<sup>[16]</sup>,且近邻数 $k$ 设置为6。实验分为:a)对细胞表型图像加入噪声,测试Kernel-SEM和QEM两种模型对分类精度的影响;b)对类别实施加噪处理,测试各自效果。

实验的第一部分,在训练样本和测试样本的属性中分别随机抽取50%加入均值为1和方差为0.1的高斯白噪声,其余的加入均值为0和方差为0.01的高斯白噪声。利用Kernel-SEM和QEM两种模型进行学习,最后利用半监督分类器获取分类性能。图2给出了Kernel-SEM和QEM两种模型在测试样本中的分类准确率,其中纵坐标为分类测试准确率(%),横坐标为有类别样本占总样本的比例(%)。

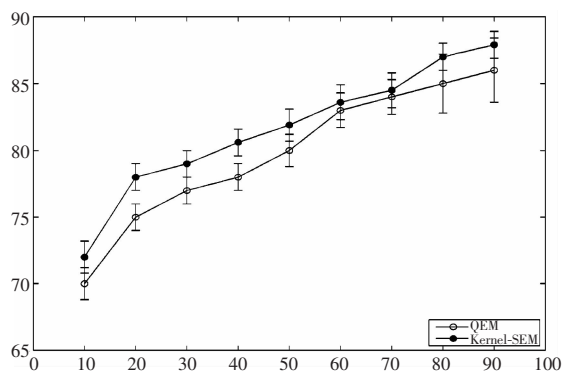


图2 2D HeLa 测试样本分类性能

从图2可知, Kernel-SEM 由于在形态差异学习模型优化过程中以迭代方式进行了降噪优化, 因此能有效抵抗加入噪声的影响, 分类器性能与文献[7]中最优性能接近; 而 QEM 在加入噪声的数据集上训练, 性能有一定程度的下降, 原因在于文献[7]中的降噪方法仅仅保留了边缘特征, 而没有考虑细胞表型图中特征的全局相似性。

实验的第二部分, 给样本类别加噪, 每类样本分别抽取10%, 并以相同的概率指定为其他类别。图3给出了对类别加噪后, 样本在经过能量模型学习后的可视化效果。

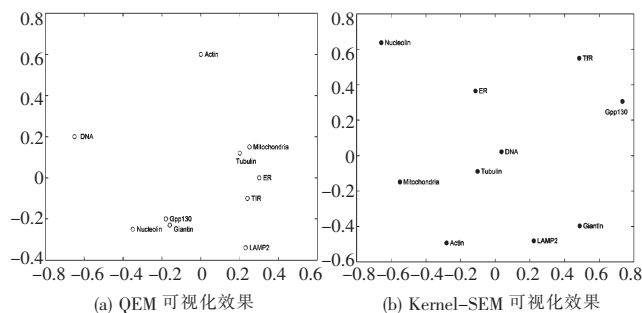


图3 能量模型分类性能可视化

由于本文形态差异学习模型在学习的过程中, 相当于在原始空间和核空间中学习一个映射, 使得在核空间中的同类样本尽量靠近, 不同类样本尽量远离, 相对于原始空间, 核空间中的特征有一个相对移动的距离。同时 NL means 的核扩展降噪形式有利于消除类别噪声的影响, 由图3可知, 各类样本在经过能量模型学习后均具有较好的聚集性, 且不同类别的样本之间分离程度较好, 有利于分类。然而图3(a) (Gpp130, Giantin) 和 (Tubulin, Mitochondria) 两对数据距离较近, 影响分类效果。而图3(b)中经过核映射之后, 各类距离较远, 能有效提升分类性能。在真实数据上的实验表明, 即使特征或者类别加入噪声, 本文的算法仍然能体现较好的鲁棒性。

### 3 结束语

本文主要考虑细胞表型图形态差异学习算法对样本噪声和类别噪声的鲁棒性以及信息能度量在提取特征时的有效性。主要工作如下:

- 信息能度量的提出, 能有效提取数据中的高阶统计量和非线性特征, 提高形态差异定量分析的灵敏性。
- 梯度上升算法保证了算法的快速求解, 同时迭代过程中兼容了 NL means 降噪模型, 能进一步提高算法的鲁棒性。
- 使用真实数据的实验证明了提出算法的有效性, 尽管取得了一些有益的结果, 但梯度算法的局部最优、噪声检测及

消除、形态差异学习算法在其他图像方面的应用需要作进一步深入研究和探讨。

### 参考文献:

- LEE Y J, JEONG S Y, KARBOWSKI M. Roles of the mammalian mitochondrial fission and fusion mediators Fis1, Drp1, and Opa1 in apoptosis [J]. *Molecular Biology of the Cell*, 2004, 15(11): 5001-5011.
- 闫海霞, 刘岩俊. 基于顺序形态学自适应边缘检测方法 [J]. *计算机应用研究*, 2011, 28(5): 1978-1980.
- LANG P, YEOW K, NICHOLS A, et al. Cellular imaging in drug discovery [J]. *Nature Reviews Drug Discovery*, 2006, 5: 343-356.
- JONES T R, CARPENTER A E, LAMPRECHT M R, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning [C] // *Proc of National Academy of Sciences*. [S. l.]: Stanford University's Highwire Press, 2009: 1862-1831.
- LOO L H, WU L F, ALTSCHULER S J. Image-based multivariate profiling of drug responses from single cells [J]. *Nature Methods*, 2007, 4(5): 445-453.
- NANNI L, LUMINI A, LIN Y S, et al. Fusion of systems for automated cell phenotype image classification [J]. *Expert Systems with Applications*, 2010, 37(2): 1556-1562.
- LIN Y S, LIN C C, TSAI Y S, et al. A spectral graph theoretic approach to quantification and calibration of collective morphological differences in cell images [J]. *Bioinformatics*, 2010, 26(12): 29-37.
- CHANG H H, MOURA J M F. Classification by Cheeger constant regularization [C] // *Proc of IEEE International Conference on Image Processing*. USA: IEEE Press, 2007: 209-212.
- BUADES A, COLL B, MOREL J. A non-local algorithm for image denoising [J]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, 2(6): 60-65.
- 徐大宏, 王润生. 基于非局部正则化的图像去噪 [J]. *计算机应用研究*, 2009, 26(12): 4830-4832.
- 陈卿, 胡绍湘. 基于边缘扩散的医学图像非线性去噪算法 [J]. *计算机应用研究*, 2011, 28(8): 3178-3180.
- 吕占强, 孙玉宝. 稀疏性正则化的图像 Laplace 去噪及 PR 算子分裂算法 [J]. *计算机应用研究*, 2011, 28(9): 3542-3544.
- CRAMMER K, SINGER Y. On the algorithmic implementation of multiclass kernel-based vector machines [J]. *Journal of Machine Learning Research*, 2001, 2(3/1): 265-292.
- 刘松华, 张军英, 许进, 等. Kernel-kNN: 基于信息能度量的核 k-最近邻算法 [J]. *自动化学报*, 2010, 36(12): 1681-1688.
- CHEBIRA A, BARBOTIN Y, JACKSON C, et al. A multiresolution approach to automated classification of protein subcellular location images [J]. *BMC Bioinformatics*, 2007, 8(1): 210.
- ZHU X J, GHAMRANI Z B, LAFFERTY J. Semi-supervised learning using Gaussian fields and harmonic functions [C] // *Proc of the 20th International Conference on Machine Learning*. Menlo Park, CA: AAAI Press, 2003: 912-919.
- TORKKOLA K. Feature extraction by non-parametric mutual information maximization [J]. *Journal of Machine Learning Research*, 2003, 2(3): 1415-1438.