

# 基于核熵成分分析的流式数据自动分群方法\*

董明利,马闪闪,张帆,潘志康

(北京信息科技大学 光电测试技术北京市重点实验室 北京 100192)

**摘要:**针对多参数流式细胞数据传统人工分群过程复杂、自动化程度不高等问题,提出了一种基于核熵成分分析(KECA)的自动分群方法。选取对瑞利(Renyi)熵具有最大贡献的特征向量作为投影方向,对数据进行特征提取;设计了一种基于余弦相似度和K-means算法的分类器,并采用一种基于向量夹角的最佳聚类数确定方法,最终获得细胞的分类标签。对实验获得的淋巴细胞免疫表型分析数据进行处理,结果表明,该方法能够实现细胞的快速、自动分群,整体分群准确率能够达到97%以上,操作简单便捷,提高了细胞分析的效率。

**关键词:**流式细胞术;自动分群;核熵成分分析;K-means算法;余弦相似度

**中图分类号:** TH773 TP391 **文献标识码:** A **国家标准学科分类代码:** 510.40

## Auto classification method of flow cytometry data based on kernel entropy component analysis

Dong Mingli, Ma Shanshan, Zhang Fan, Pan Zhikang

(Beijing Key Laboratory for Optoelectronics Measurement Technology, Beijing Information Science and Technology University, Beijing 100192, China)

**Abstract:** The traditional clustering process of multi-parametric flow cytometry data analysis is complicated, non-automated and time-consuming. To overcome this limitation, an automatic clustering method based on kernel entropy analysis (KECA) is proposed. The feature vector with the greatest contribution to the Renyi entropy is selected as the projection direction to carry out the feature extraction. A classifier based on cosine similarity and the K-means algorithm is designed to get the label of each cell, and a method for determining the optimal number of clusters based on the angle of vectors is adopted. Experimental data of peripheral blood lymphocyte is processed, and the results indicate that the proposed method can realize automatically clustering with simple operation and the overall accuracy rate of clustering can reach 97%, which can improve the efficiency of cell analysis.

**Keywords:** flow cytometry; automatic clustering; kernel entropy component analysis (KECA); K-means algorithm; cosine similarity

## 1 引言

流式细胞术是一种能够对悬浮的细胞或者其他生物粒子进行定量分析和分选的检测技术,被广泛应用于临床医学和基础医学等研究领域<sup>[1-2]</sup>。随着流式细胞术的发展,流式细胞仪能够检测的参数成倍增加,对多参数流式细胞数据进行快速、准确的分析是提高临床诊断效率的关键。

经荧光染料标记的细胞在激发光源的照射下能够发出表征细胞物理特性的前向散射光(forward scatter, FS)、侧向散射光(side scatter, SS)信号以及表征化学特性的荧光信号(fluorescence, FL)<sup>[3]</sup>。传统的流式数据分析方法是根据细胞的散射光或荧光的特性,通过人工设门的方式在一维峰图或二维散点图上选定固定区域内的细胞进行分析<sup>[4-5]</sup>。该方法存在以下缺陷:1)分析结果可重复性差,常受到操作者的专业技能及操作经验的影响;2)分析过程繁琐,耗时长;3)不能同时分析流式细胞数

据的多个维度。传统的人工分析方法已不能满足流式细胞仪临床快速分析诊断的需求,因此提出一种能够对流式细胞数据进行准确、快速分析的方法十分必要。

近年来,一些聚类算法如 K-means 算法、高斯混合模型等被提出应用于流式细胞数据的聚类分析中<sup>[6-8]</sup>。Zeng Q. T. 等人<sup>[9]</sup>由筛选器从多个维度的直方图中提取出峰值点,通过对比类群的中心位置与峰值位置是否匹配,确定细胞类群的数目;Lo K 等人<sup>[10]</sup>提出一种基于多元  $t$  混合模型的聚类方法,并结合 Box-Cox 变换对流式数据进行处理,灵活性较好;Zare H 等人<sup>[2]</sup>通过添加信息保护采样过程,提出了修正谱聚类算法,首次应用谱聚类算法实现了流式数据自动分析;Sugar I. P. 等人<sup>[11]</sup>开发了一种基于渗流理论的无监督密度轮廓聚类算法,实现了数据的快速分析;Ge Y 等人<sup>[12]</sup>结合了有限元混合模型与直方图,利用混合模型密度函数确定类群数目;Aghaeipour N 等人<sup>[13]</sup>采用变点检测算法确定类群数目,实现了细胞类群的快速识别;王先文等人<sup>[14]</sup>提出了基于偏斜  $t$  分布的混合模型聚类方法,通过有限混合模型对数据进行统计分析,实现了细胞的自动设门。以上研究中,大多数方法的最终目的是获取细胞事件的分类标签,想要获取较好的细胞分群结果,仍需要操作者根据相关背景知识和样本染色策略手动选择散点图的坐标轴参数,未实现真正的自动分群。

为了简化人工选取坐标轴绘制散点图的过程,普度大学 Grégori G 等人<sup>[15]</sup>采用主成分分析法(principal component analysis, PCA)处理光谱流式细胞仪获得的多维流式数据,以贡献度最高的两个主成分作为横、纵坐标自动绘制散点图,实现了光谱流式数据的自动分群。PCA 是一种常用的线性变换方法,可有效提取特征之间的线性关系,但不能解决非线性问题。

Scholkopf B 等人<sup>[16]</sup>提出的核主成分分析(kernel principal component analysis, KPCA)是 PCA 的非线性扩展方法,它通过映射函数把原始数据映射到高维特征空间上,采用非线性方法抽取核主成分,使数据在原始空间的线性不可分变成高维空间的线性可分。核熵成分分析(kernel entropy component analysis, KECA)于 2010 年被 Jenssen R 提出,是一种新的特征提取和降维方法<sup>[17]</sup>。同 KPCA 相似,该算法也是将原始数据投影至高维特征空间,其不同之处是 KECA 选取对瑞利(Renyi)熵具有最大贡献的特征向量作为投影方向,而不是特征值较大的主元方向<sup>[17-18]</sup>。目前,KECA 被广泛应用于故障检测及图像分类等方面<sup>[19-22]</sup>。本文将 KECA 应用于多参数流式细胞数据分析中,对多参数流式细胞数据进行特征提取和降维处理,同时设计了一种基于余弦相似度和 K-means 算法的分类器对变换后的数据进行聚类分析,通过自动设置散点图的横、纵坐标轴,实现细胞的自动分群。

## 2 KECA 的流式数据分群方法

### 2.1 KECA

信息熵是由 Shannon C. E. 提出的,它是数据总体信息不确定性的量度,信息熵越大,信息量越多<sup>[18]</sup>。KECA 将核函数学习方法与信息熵进行了有机结合,对核矩阵进行特征分解,选取对信息熵具有较大贡献的特征向量作为投影方向进行投影,得到变换后的数据,KECA 分析过程如下<sup>[17-20]</sup>。

设原始数据样本矩阵形式为  $\mathbf{X} = x_1, x_2, \dots, x_N$ , 概率密度函数是  $\rho(\mathbf{x})$ , 则二次 Renyi 熵表达式为:

$$H(\rho) = -\log \int \rho^2(\mathbf{x}) d\mathbf{x} \quad (1)$$

对数函数为单调函数,令  $V(\rho) = \int \rho^2(\mathbf{x}) d\mathbf{x}$ , 采用 Parzen 窗密度估计方法对  $V(\rho)$  进行估计,  $k_\sigma(\mathbf{x}, \mathbf{x}_i)$  为 Mercer 核函数,  $\sigma$  为核函数的宽度参数, Parzen 窗为:

$$\hat{\rho}(\mathbf{x}) = \frac{1}{N} \sum_{x_i \in D} k_\sigma(\mathbf{x}, \mathbf{x}_i) \quad (2)$$

则二次 Renyi 熵估计可表示为:

$$\hat{V}(\rho) = \frac{1}{N} \sum_{x_i \in D} \hat{\rho}(\mathbf{x}_i) = \frac{1}{N} \sum_{x_i \in D} \frac{1}{N} \sum_{x_j \in D} k_\sigma(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{N^2} \mathbf{1}^T \mathbf{K} \mathbf{1} \quad (3)$$

式中:  $\mathbf{1}$  为  $N \times 1$  的单位向量,  $\mathbf{K}$  为  $N \times N$  的样本核矩阵。对核矩阵  $\mathbf{K}$  进行特征分解:  $\mathbf{K} = \mathbf{E} \mathbf{D} \mathbf{E}^T$ , 其中  $\mathbf{D}$  是特征值矩阵,  $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ ,  $\mathbf{E}$  是特征向量矩阵,  $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$ , 可知二次 Renyi 熵估计基于核矩阵的特征值和特征向量的表达式为:

$$\hat{V}(\rho) = \frac{1}{N^2} \sum_{i=1}^N (\sqrt{\lambda_i} \mathbf{e}_i^T \mathbf{1})^2 \quad (4)$$

由式(4)可知,二次 Renyi 熵估计不仅与特征值有关,且与特征值所对应的特征向量各元素之和有关。

多参数流式细胞数据可以表示为  $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]^T$ , 其中  $M$  为细胞的总个数,通过非线性映射  $\varphi: \mathbf{x}_i \rightarrow \varphi(\mathbf{x}_i)$ , 将原始多参数流式细胞数据映射到特征空间,则  $\Phi(\mathbf{x}) = [\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_M)]$ 。优先选取对  $\hat{V}(\rho)$  具有较大贡献的前  $l$  个特征值及其对应的特征向量,产生 KECA 映射  $\Phi_{\text{eca}} = \mathbf{D}_l^{1/2} \mathbf{E}_l^T$ , 得到新的特征空间中的数据。

### 2.2 基于余弦相似度和 K-means 的分类器

传统的 K-means 算法是典型的基于距离的聚类算法,采用欧氏距离作为相似性的评价指标,距离越近,其相似度越大。由于流式数据经 KECA 方法转换后在特征空间有较好的簇可分离性,不同的类群关于原点都保持着一定的角度,根据该特点,设计了基于余弦相似度和 K-means 算法的分类器,对特征提取后的流式数据进行分类。

余弦相似度也称为余弦距离,与欧氏距离不同,它是利用空间中两个向量的夹角余弦值作为衡量两个个体之间差异大小的度量。将每个细胞的流式数据看作为一个多维的特征向量,通过计算两个向量之间的夹角余弦值来评估其相似度。

假设两个细胞均包含  $m$  维个特征,利用特征向量表示分别为  $\mathbf{a} = (a_1, a_2, \dots, a_m)$ 、 $\mathbf{b} = (b_1, b_2, \dots, b_m)$ ,则其夹角余弦值可以表示如下:

$$\cos\theta = \frac{\sum_{i=1}^m (a_i \times b_i)}{\sqrt{\sum_{i=1}^m a_i^2} \times \sqrt{\sum_{i=1}^m b_i^2}} \quad (5)$$

如图1所示,若向量  $\mathbf{a}$ 、 $\mathbf{b}$  的方向一致,则二者之间的夹角接近于  $0^\circ$ ,夹角余弦值越接近于1,两个向量相似性越高;若两个特征向量的夹角很大,其夹角余弦值越小,二者之间的相似性也越低。

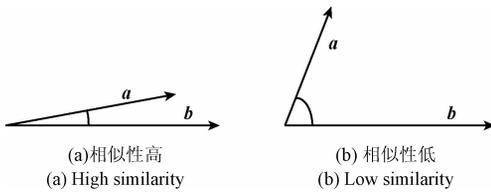


图1 特征向量夹角

Fig.1 Feature vector angle

借助三维坐标系对欧氏距离  $\text{Dist}(\mathbf{a}, \mathbf{b})$  和余弦距离  $\cos\theta$  进行对比,如图2所示。

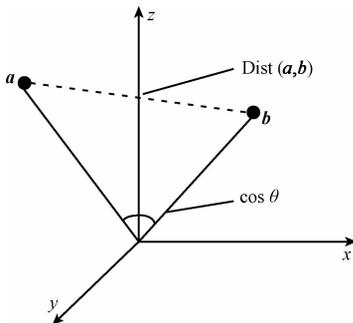


图2 欧氏距离与余弦距离对比

Fig.2 Comparison of the Euclidean distance and the cosine distance

与欧氏距离相比,余弦距离更侧重于两个向量在方向上的差异,而对绝对数值并不敏感,因此可修正样本特征数值中可能存在的度量标准不统一的问题。由于夹角余弦值的范围为  $[-1, 1]$ ,余弦值越大,说明两个特征向量的方向更一致,相似度也越高。以余弦相似度作为评

价标准,结合 K-means 算法获得细胞的分类标签。

基于余弦相似度和 K-means 的分类器的具体步骤如下:

- 1) 确定聚类个数  $C$ ;
- 2) 计算样本数据特征向量之间的夹角余弦值,获得夹角余弦矩阵;
- 3) 选取夹角余弦值最小的两个向量作为前两个初始聚类中心;
- 4) 选取与前两个初始聚类中心之间夹角余弦值最小的数据向量作为第3个初始聚类中心;
- 5) 依次确定  $C$  个初始聚类中心;
- 6) 根据余弦相似度大小,将其它数据样本向量初步划分到以  $C$  个聚类中心为代表的类群中;
- 7) 更新每个类群的聚类中心;
- 8) 重复步骤6)~7),直到各类群中心不再发生改变,则停止,获得样本数据的分类标签。

### 2.3 最佳聚类数目的确定

针对 K-means 算法无法确定聚类数目的问题,本文利用向量夹角法对原始数据进行处理,以确定聚类数的初始选取范围。通过运行 K-means 算法得到不同聚类数下的聚类结果,根据聚类指标对聚类结果进行评估,从而确定最佳聚类数目。

向量夹角法思路如下:将经过 KECA 处理后的流式数据的每一行看作是一个特征向量,计算每个特征向量与其他向量之间的夹角,对其进行升序排列后计算相邻夹角之间的差值,得到向量夹角变化数组。由于不同类群之间的向量夹角大于同种类群之间的向量夹角,因此,属于同种类群的特征向量所对应的夹角差值数组的分布趋势近似。分别绘制每个特征向量对应的夹角差值变化曲线,根据曲线上不同位置出现的波峰总个数确定聚类数目的初始范围上限。

聚类的目标是将样本分为若干个类群,并保证类群内的样本尽可能密集,不同类群之间尽可能离散。聚类指标是衡量聚类结果是否最优的标准,通过引入类内夹角余弦平均值和类间夹角余弦平均值,构建一种新的聚类指标,确定最佳聚类数。

类内夹角余弦平均值为:

$$L_i = \frac{1}{c} \sum_{i=1}^c \sum_{x_i, x_{i'} \in C_i} \cos \angle(x_i, x_{i'}) \quad (6)$$

式中:  $i = 1, 2, \dots, C$ ,  $C$  为聚类数,  $C_i$  为第  $i$  个类群中的元素,  $M_i$  为第  $i$  个类群的元素总数。

类间夹角余弦平均值为:

$$L_o = \frac{1}{M^2 - \sum_{i=1}^c M_i^2} \sum_{i \neq j} \sum_{x_i, x_j \in C_i, C_j} \cos \angle(x_i, x_j) \quad (7)$$

式中:  $i, j = 1, 2, \dots, C$ 。

根据聚类的最佳效果是类内夹角余弦值越大越好,类间夹角余弦值越小越好,利用聚类指标  $S = \max(L_i - L_o)$  来衡量聚类效果的好坏,聚类指标取得的最大值对应最佳聚类数  $C$ 。

### 3 实验与分析

为了验证基于 KECA 的流式细胞数据自动分群方法的有效性,将该方法应用于淋巴细胞免疫表型分析中。实验数据来自于某健康志愿者的外周血淋巴细胞样本,其包含 4 698 个细胞以及 4 种标志分子即 CD3、CD45、CD19 和 CD56,实验仪器为美国 BD 公司 (Becton, Dickinson and Company) 的 FACSCalibur 流式细胞仪。该样本流式数据包括 10 个参数,分别为脉冲高度 (FITC-H, PE-H, PerCP-H, APC-H)、脉冲面积 (FSC-A, SSC-A, FITC-A, PE-A, PerCP-A, APC-A)。

首先绘制各样本数据的特征向量对应的夹角变化曲线,根据出现的波峰总个数,确定聚类数初始范围为 [2, 5],以聚类数为自变量绘制聚类指标取值如图 3 所示。

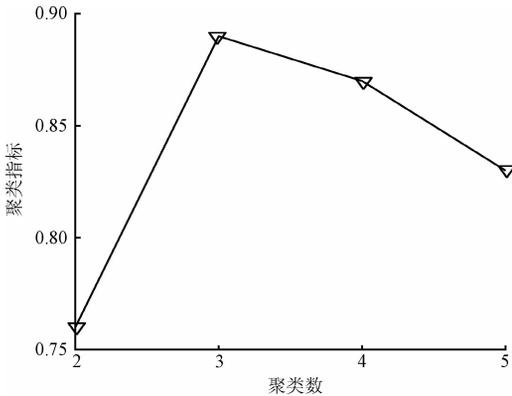


图 3 聚类数与指标之间的关系

Fig. 3 Relationship between the number of clusters and the index

根据聚类数与聚类指标之间的关系图,可以看出当聚类数  $C$  为 3 时,聚类指标的值最大,因此确定样本数据的最佳聚类数目为 3。

最佳聚类数目确定后,分别采用 KPCA、KECA 对实验数据进行特征提取,KPCA 处理得到的前 6 个核主成分及其所占的累计方差贡献率和 KECA 处理得到的核熵主成分及其对 Renyi 熵的累计贡献率如表 1 所示。

由表 1 可以看出,经 KPCA 和 KECA 处理后得到的数据均能很好地保留原始数据的特征信息。对比 KPCA 和 KECA 特征提取的结果,若均选取前两个主成分,则 KECA 处理后的数据保留原始数据在核特征空间均值向量的比值能够达到 97.98%,而 KPCA 处理得到的前两个

主成分的累积方差贡献率仅能够达到 79.54%。两种方法的不同之处在于:KPCA 以数据在特征空间中的方差大小为标准,根据特征值的大小依次选取主成分;而 KECA 则以数据的 Renyi 熵值大小为标准,选取对 Renyi 熵贡献率较大的主元方向作为投影方向,综合考虑了特征值的大小与其相应的特征向量元素之和两个方面。

表 1 KPCA 和 KECA 特征提取

Table 1 Feature extraction of KPCA and KECA

KPCA		KECA	
特征值	累计方差/%	特征值	累积 Renyi 熵/%
1.741 3	61.69	1.741 3	94.39
0.503 6	79.54	0.283 1	97.98
0.283 1	89.57	0.503 6	99.25
0.183 5	96.07	0.183 5	99.99
0.070 8	98.57	0.025 0	99.99
0.025 0	99.46	0.070 8	100%

为了观察本文方法得到的流式数据自动分群效果,分别将 KPCA 变换后的特征值最大的前两个核主成分和 KECA 变换后对 Renyi 熵贡献最大的前两个核熵主成分设置为横、纵坐标参数,绘制的二维散点图,并利用本文提出的分类算法对数据进行聚类,进行设门,得到 KPCA、KECA 分群结果如图 4 所示。

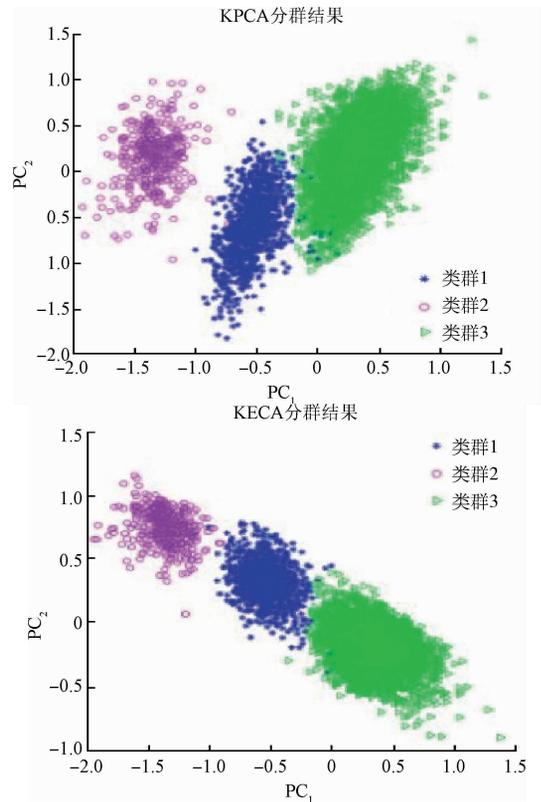


图 4 KPCA 和 KECA 分群结果

Fig. 4 Clustering results of KPCA and KECA

由图4可以看出,KECA处理得到的二维散点图中分群结果更好。通过对比KECA、KPCA的处理过程可知,在KPCA分析中,所选取的特征值最大的前两个核主成分分别为第1、2主成分( $PC_1, PC_2$ ),而在KECA分析中,所选取的对Renyi熵贡献最大的前两个核熵主成分分别对应为KPCA分析中的第1、3主成分( $PC_1, PC_3$ )。为了对比KPCA和KECA所选取主成分的不同,分别绘制核主成分 $PC_1 \sim PC_3$ 的核密度估计如图5所示。

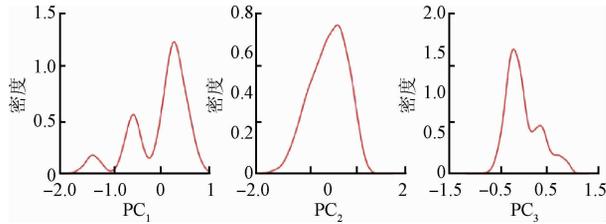


图5 核主成分变量 $PC_1 \sim PC_3$ 的核密度估计图

Fig. 5 Kernel density estimation of  $PC_1 \sim PC_3$

观察图5,由 $PC_1$ 的核密度估计图可直观观察3种类群的分布情况和密度中心趋势,同时根据波峰个数可判断细胞类群数,验证了最佳聚类数目确定方法的准确性;与 $PC_1, PC_3$ 相比, $PC_2$ 提供的数据分布特征信息最少。结合核密度估计图和二维散点图进行分析可知,若自动设置前两个主变量作为横、纵坐标进行细胞分群分析,KECA更能准确地提取和保留原始数据的特征信息及不同类群数据之间的差异性。

利用本文设计的分类器对经KECA特征提取后的数据进行聚类分析,以传统人工分群结果作为参考标准,选择不同的核参数 $\sigma$ 时,得到的分群准确率如图6所示。结果表明,当 $\sigma > 1.8$ 后,淋巴细胞实验数据的整体分群准确率一直维持在97%以上。 $\sigma$ 最优时,3种细胞亚群的分群准确率分别为99.15%、95.59%和99.47%,平均分群准确率为98.07%,满足实验分析的要求。

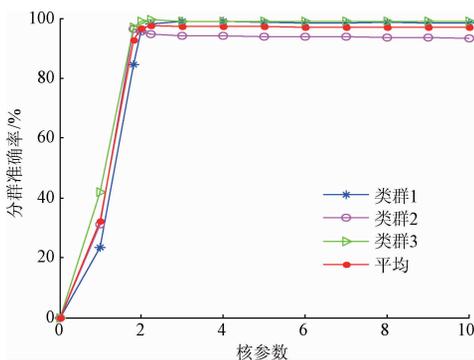


图6 分群准确率与 $\sigma$ 之间的关系

Fig. 6 Clustering accuracy as a function of  $\sigma$

## 4 结 论

提出了一种基于KECA的流式细胞数据自动分群方法。由KECA对流式细胞数据进行非线性特征提取,转换后的数据不仅降低了维度,而且在特征空间中最大程度地呈现了良好的簇可分离性。根据处理后的数据特点,设计了余弦相似度和K-means聚类算法相结合的分类器对数据进行聚类,实现细胞的自动设门。以淋巴细胞免疫表型分析的实验数据为处理对象,结果表明,本文方法通过自动设置前两个核熵主成分为散点图坐标轴,实现了细胞的自动分群;无需人工多次选取散点图坐标轴参数,降低了对操作者的专业要求,同时克服了传统人工分析方法中步骤繁琐、可重复性差、效率低等不足。以人工理想分群为参考标准,当 $\sigma > 1.8$ 时,本文方法所获得的整体分群准确率能够稳定维持在97%以上,可靠性高。本文方法大大提高了流式细胞数据的分群效率,在流式细胞仪快速、自动分析方面有较好的应用前景。

## 参考文献

- [1] PEDREIRA C E, COSTA E S, LECREVISSE Q, et al. Overview of clinical flow cytometry data analysis: Recent advances and future challenges [J]. Trends in Biotechnology, 2013, 31(7):415-425.
- [2] ZARE H, SHOOSHTARI P, GUPTA A, et al. Data reduction for spectral clustering to analyze high throughput flow cytometry data[J]. BMC Bioinformatics, 2009, 11(1):1-16.
- [3] 张文昌, 祝连庆, 娄小平, 等. 基于灰色预测恢复算法的流式细胞仪多参数提取[J]. 仪器仪表学报, 2015, 36(7):1660-1665.  
ZHANG W CH, ZHU L Q, LOU X P, et al. Multi-parameter extraction of flow cytometer based on grey prediction recovery algorithm [J]. Chinese Journal of Scientific Instrument, 2015, 36(7):1660-1665.
- [4] ROBINSON J P. Computational analysis of high-throughput flow cytometry data [J]. Expert opinion on drug discovery, 2012, 7(8):679-693.
- [5] BASHASHATI A, BRINKMAN R R. A survey of flow cytometry data analysis methods [C]. Advance in Bioinformatics, 2009:584603.
- [6] MOSMANN T R, NAIM I, DATTA S, et al. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 1: Algorithm design [J]. Cytometry Part A, 2014, 85(5):408-421.
- [7] WILKINS M F, HARDY S A, BODDY L, et al. Comparison of five clustering algorithms to classify

- phytoplankton from flow cytometry data [J]. Cytometry Part A, 2001, 44(3):210-217.
- [8] CHAN C, FENG F, OTTINGER J, et al. Statistical mixture modeling for cell subtype identification in flow cytometry [J]. Cytometry Part A the Journal of the International Society for Analytical Cytology, 2008, 73(8):693-701.
- [9] ZENG Q T, PRATT J P, PAK J, et al. Feature-guided clustering of multi-dimensional flow cytometry datasets [J]. Journal of Biomedical Informatics, 2007, 40(3):325-331.
- [10] LO K, BRINKMAN R R, GOTTARDO R. Automated gating of flow cytometry data via robust model-based clustering [J]. Cytometry Part A, 2008, 73A(4):321-332.
- [11] SUGAR I P, SEALFON S C. Misty mountain clustering: application to fast unsupervised flow cytometry gating [J]. Bmc Bioinformatics, 2010, 11(3):2277-2289.
- [12] GE Y, SEALFON S C. FlowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding [J]. Bioinformatics, 2012, 28(15):2052-2058.
- [13] AGHAEPOUR N, NIKOLIC R, HOOS H H, et al. Rapid cell population identification in flow cytometry data [J]. Cytometry Part A, 2011, 79(1):6-13.
- [14] 王先文, 陈锋, 程智, 等. 基于偏斜  $t$  混合模型的流式数据自动聚类方法研究 [J]. 电子学报, 2014, 42(12):2527-2535.  
WANG X W, CHEN F, CHENG ZH, et al. Auto clustering method study of flow cytometry data based on skew  $t$ -mixture models [J]. Acta Electronica Sinica, 2014, 42(12):2527-2535.
- [15] GREGORI G, PATSEKIN V, RAJWA B, et al. Hyperspectral cytometry at the single-cell level using a 32-channel photodetector [J]. Cytometry Part A, 2012, 81(1):35-44.
- [16] SCHOLKOPF B, SMOLA A, MULLER K. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5):1299-1319.
- [17] JENSSEN R. Kernel entropy component analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(5):847-860.
- [18] SHI J, JIANG Q, MAO R, et al. FR-KECA: Fuzzy robust kernel entropy component analysis [J]. Neurocomputing, 2015, 149(2):1415-1423.
- [19] 王瀛, 郭雷, 梁楠. 基于核熵成分分析的高光谱遥感图像分类算法 [J]. 吉林大学学报:工学版, 2012, 42(6):1597-1601.
- WANG Y, GUO L, LIANG N. Classification algorithm of hyperspectral images based on kernel entropy analysis [J]. Journal of Jilin University: Engineering and Technology Edition, 2012, 42(6):1597-1601.
- [20] 张朝龙, 何恰刚, 袁莉芬, 等. 基于核熵成分分析的模拟电路早期故障诊断方法 [J]. 仪器仪表学报, 2015, 36(3):675-684.  
ZHANG CH L, HE Y G, YUAN L F, et al. Approach for analog circuit incipient fault diagnosis based on KECA [J]. Chinese Journal of Scientific Instrument, 2015, 36(3):675-684.
- [21] 齐咏生, 张海利, 高学金, 等. 基于 KECA 的化工过程故障监测新方法 [J]. 化工学报, 2016, 67(3):1063-1069.  
QI Y SH, ZHANG H L, GAO X J, et al. Novel fault monitoring strategy for chemical process based on KECA [J]. CIESC Journal, 2016, 67(3):1063-1069.
- [22] 何飞, 徐金梧, 梁治国, 等. 基于核熵成分分析的热轧带钢自适应聚类分析 [J]. 中南大学学报:自然科学版, 2012, 43(5):1732-1738.  
HE F, XU J W, LIANG ZH G, et al. Hot rolled strip state clustering based on kernel entropy component analysis [J]. Journal of Central South University: Science and Technology, 2012, 43(5):1732-1738.

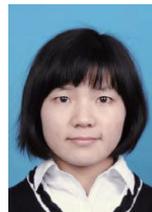
## 作者简介



**董明利** (通讯作者), 1989 年于合肥工业大学获得硕士学位, 2009 年于北京理工大学获得博士学位, 现为北京信息科技大学教授, 主要研究方向为生物医学检测技术与仪器、视觉测量技术。

E-mail: dongml@sina.com

**Dong Mingli** (Corresponding author) received her M. Sc. degree in 1989 from Hefei University of Technology, and her Ph. D. degree in 2009 from Beijing Institute of Technology. Now she is a professor in Beijing Information Science and Technology University. Her main research interests include biomedical detection technology and instrument, vision measurement technology.



**马闪闪**, 2014 年于北京信息科技大学获得学士学位, 现为北京信息科技大学硕士研究生, 主要研究方向为生物医学检测技术与仪器。

E-mail: mashan1028@163.com

**Ma Shanshan** received her B. Sc. degree in 2014 from Beijing Information Science and Technology University. Now she is a master student in Beijing Information Science and Technology University. Her main research interest is biomedical detection technology and instrument.