

基于 Laplacian 中心性的密度聚类算法

杨旭华¹ 朱钦鹏¹ 童长飞²

(浙江工业大学计算机科学与技术学院 杭州 310023)¹

(温州大学计算机科学与工程系 浙江 温州 325035)²

摘要 聚类分析是一种重要的数据挖掘工具,可以衡量不同数据之间的相似性,并把它们分到不同的类别中,在模式识别、经济学和生物学等领域有着广泛的应用。文中提出了一种新的聚类算法。首先,把待分类的数据集转换成一个加权的完全图,每个数据点为一个节点,两个数据点之间的距离为相应两个节点之间边的权值。然后,用 Laplacian 中心性来计算和评价该网络每个节点的局部重要性,聚类中心为局部的密度中心,它具有比周围的邻居节点更高的 Laplacian 中心性,并且与具有更高 Laplacian 中心性的节点之间的距离也较大。新算法是一种真正的无参数聚类方法,不需要任何先验参数便可以自动地对数据集进行分类。在 6 种数据集中将其与 9 种知名聚类算法做了对比,结果显示该算法具有良好的聚类效果。

关键词 加权完全图, Laplacian 中心性, 密度聚类

中图法分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.01.051

Density Clustering Algorithm Based on Laplacian Centrality

YANG Xu-hua¹ ZHU Qin-peng¹ TONG Chang-fei²

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)¹

(Department of Computer Science and Engineering, Wenzhou University, Wenzhou, Zhejiang 325035, China)²

Abstract As an important tool of data mining, clustering analysis can measure similarity between the different data and classify them into different categories. It is wisely applied in pattern recognition, economics, biology and so on. In this paper, a new clustering algorithm was proposed. Firstly, dataset to be classified is converted into a weighted complete graph. Data point is a node and the distance between two data points is used as weight of side between these two data points. Secondly, local importance of each node in the network is calculated and evaluated by Laplacian centrality. The cluster center has higher Laplacian centrality than surrounding neighbor nodes and the node with higher Laplacian centrality has larger distance. Finally, the algorithm is a real parameter-free clustering method, which can classify the dataset automatically without any priori parameters. In this article, the new algorithm was compared with 9 famous clustering algorithms in 6 datasets. Experimental results show that the proposed algorithm has good clustering performance.

Keywords Weighted complete graph, Laplacian centrality, Density clustering

1 引言

机器学习^[1]是人工智能领域中一个十分活跃的重要分支,其研究目标主要是让计算机模拟或实现人类的学习行为,以获取新的知识或技能。聚类算法^[2]属于机器学习中的无监督学习算法,可以提取数据内在的隐含结构,将具有相似性属性的数据点聚成一个类簇,类簇内部的数据点具有较大的相似性,而不同类簇的数据点之间的相似性比较小。聚类算法被广泛应用于计算机和其他学科各个领域^[3-6]。

研究者们对聚类算法已经做出了很多有益的工作,比如基于划分的聚类算法:K-means^[7]是基于一个优化的目标函数,通过反复迭代来降低目标函数的差异值,当目标函数逐渐收敛时,划分成 k 个类簇。该算法的优点是快速高效;缺点是需要先验指定类簇数目 k 或类簇中心,聚类的鲁棒性不高。因此后来的研究者对其进行了改进,如提出 K-medoids^[8], K-means++^[9] 等。基于密度的聚类算法:DBSCAN^[10]算法是基于密度的空间聚类算法,该算法采用空间索引技术来搜索对象的邻域,引入了“核心对象”和“密度可达”等概念,从核心

到稿日期:2016-12-01 返修日期:2017-03-23 本文受国家自然科学基金(61374152),浙江省自然科学基金(LY17F030016, LQ13G010007)资助。

杨旭华(1971-),男,博士,教授,主要研究方向为网络科学、智能交通、机器学习, E-mail: xhyang@zjut.edu.cn(通信作者);朱钦鹏(1991-),男,硕士生,主要研究方向为网络科学、聚类算法、社团划分, E-mail: sh2zqp@163.com;童长飞(1980-),男,博士,副教授,主要研究方向为数据挖掘、智能公交信息处理、物联网应用技术, E-mail: changfei@wzu.edu.cn。

对象出发,把所有密度可达的对象组成一个类簇。该算法的优点是聚类速度快且能够有效处理噪声点和发现任意形状的空间类簇;缺点是内存开销较大,且当空间类簇的密度不均匀、类簇间距离相差很大时类簇的划分效果也较差。基于层次的聚类算法——BIRCH^[11]算法,利用树结构对数据集进行处理,叶结点存储一个聚类,用中心和半径表示,顺序处理每一个对象,并把它划分到距离最近的结点,该算法也可以作为其他聚类算法的预处理过程。其优点是不需要大量递归运算,速度快,数据所需存储空间少;缺点是对输入数据的先后顺序敏感,鲁棒性不高。基于网格的聚类算法:CLIQUE^[12]算法是一种结合了网格和密度的聚类算法,利用自顶向上方法求出各个子空间的聚类单元。该算法的优点是对数据的输入顺序不敏感;缺点是空间效率和时间效率都较低,需要输入的参数具有不确定性。另外,还有很多其他聚类算法,如基于自组织神经网络和基于统计方法的聚类算法等。

除上述经典聚类算法外,一种新的基于 Delta-Density 的聚类算法(DDC)于 Science 发表^[13],并引起了学者们的广泛关注,其核心思想是巧妙地提出和利用了密度和一个基于距离的新指标 δ 去快速发现高密度数据点峰值,具有高密度和高 δ 值的数据点被确定为类簇中心;然后把去除类簇中心的数据点逐步分配到已经确定的类簇中心,直至完成类簇的划分。实验结果表明其具有良好的聚类效果;但是该算法也有不足之处,如 DDC 算法需要事先给定一个经验参数截断距离,该参数对聚类性能影响巨大,但没有一个能有效估计其合适数值的方法。针对这些问题,也有不少的改进算法被提出^[14-16]。

鉴于经典聚类算法和新型 DDC 算法存在上述不足,本文提出一种基于 Laplacian 中心性的密度峰值聚类算法(LD-DC)。首先把原始数据集转化为一个加权的完全图,然后引入 Laplacian 中心性来计算和评价加权网络^[17-20]中每个数据点的重要性,整个过程不需要人为输入任何经验参数值,完全自动地对数据集进行分类,实现了真正的无参数聚类。为了评价 LDDC 算法的性能,选取若干经典数据集和评价指标,与 DDC 以及一系列聚类算法进行了对比分析,实验结果表明 LDDC 具有令人满意的聚类效果。

本文第 2 节详细介绍了 LDDC 算法;第 3 节选取了若干具有代表性的数据集和评价指标对本文算法进行实验仿真并对结果进行分析;最后给出了本文的结论并对未来工作进行展望。

2 基于 Laplacian 中心性的密度峰值聚类算法

首先,定义一些指标和简记符号:待聚类数据集 $D = \{v_i\}, i=1, \dots, N, v_i$ 为数据集 D 中的任意数据点,数据点 v_i 和 v_j 之间的距离值为 d_{ij} ;其次,为了挖掘数据集内的类簇中心,需要对每个数据点定义两个指标:数据点 v_i 的基于 Laplacian 中心性的局部重要性指标 c_i ,数据点 v_i 到其他有更高 Laplacian 中心性的节点的最小距离值 δ_i 。其中,数据点之间距离的计算方法如下:

1)如果数据集 D 中的每个数据点均有其坐标值,则任意两个数据点的距离是指二者之间的欧氏距离。

2)如果数据集 D 中不是每个数据点都有坐标值,而是表示数据点之间的关系,如本文 3.1 节中的空手道俱乐部网络数据集是一个邻接矩阵表示的网络,网络中每一个节点是一个空手道成员,如果两个空手道成员之间有联络,则对应的两个节点之间有连边,且此时数据集 D 中的任意两个数据点 v_i 和 v_j 之间的距离 d_{ij} 为两个数据点之间的路径长度。

2.1 算法原理

首先将待聚类数据集转换成一个加权完全图,数据集中每个数据点为一个节点,两个数据点之间的距离作为相应两个节点之间边的权值。用 Laplacian 中心性来计算和评价该加权完全图中每个节点的局部重要性。

LDDC 算法的核心是对每个数据点都基于这样的一些假定:类簇中心的局部重要性要高于其邻居的局部重要性,类簇中心与具有更高局部重要性的数据点总会有相对较大的距离。计算加权完全图中每个节点的 Laplacian 中心性,对于一个数据点 v_i ,用 Laplacian 中心性作为评价其局部重要性的指标值 c_i ,将该数据点 c_i 与其他数据点的 Laplacian 中心性数值作比较,得到所有局部重要性比自己高的数据点集合,并计算自己与该集合中每个数据点的距离的最小值 δ_i 。如果一个数据点 v_i 的 c_i 和 δ_i 相较于其他普通的数据点显得异常大,那么该数据点为类簇中心,遍历加权完全图,找到所有具有这种特征的数据点作为类簇中心。对其余的数据点,根据其 c 和 δ 值将它们归属到距离最近的类簇中心所属的类簇中,具体过程如 2.4 节所述。LDDC 算法的整体描述如下:

1)预处理待分类数据集,计算任意两个数据点之间的距离,使待分类数据集转化成一个加权完全图,其中任意两个数据点之间边的权值就是这两点之间的距离;

2)计算加权完全图中任意数据点 v_i 的 Laplacian 中心性 c_i 和最小距离值 δ_i ;

3)对于所有数据点,分别以 Laplacian 中心性和最小距离值为横纵坐标做出决策图,由决策图自动地选择出坐标轴上右上角“异常大”的数据点,并将其作为数据集的类簇中心;

4)分配剩余不是类簇中心的数据点到相应的最近邻类簇中心,完成聚类。

2.2 基于 Laplacian 中心性的局部重要性指标 c_i

本文采用 Laplacian 中心性^[21]来计算加权网络节点的局部重要性,它通过计算一个数据点从网络中被移除后网络的 Laplacian 能量 $E_L(G) = \sum_i \gamma_i^2$ 的下降程度,来判断该点在网络中的重要性,其中 G 为加权网络, γ_i 为 G 的 Laplacian 矩阵的特征根;对于任意一个节点而言,它是一种介于局部和全局邻居节点信息的评价节点中心性的方法,比那些基于直接邻居的评价节点中心性的方法具有更多的结构信息,因而具有更高的准确性,可以用来计算节点在加权网络中的重要性。计算过程如下:

用 G 表示数据集的加权完全耦合网络(本文主要考虑无向无自环的网络), $W(G)$ 表示加权网络 G 的权值矩阵, $w_{i,j}$ 为

数据点 v_i 与 v_j 之间的边权值(边权值为两数据点之间的距离),如下所示:

$$W(G) = \begin{pmatrix} 0 & \omega_{1,2} & \cdots & \omega_{1,n} \\ \omega_{2,1} & \ddots & & \omega_{2,n} \\ \vdots & & \ddots & \vdots \\ \omega_{n,1} & \omega_{n,2} & \cdots & 0 \end{pmatrix} \quad (1)$$

其中, $X(G)$ 表示每个数据点到其他所有数据点的权重的和所组成的对角矩阵,距离该数据点越近,对该数据点的影响力越强,反之对该数据点的影响力就越弱。 $x_i = \sum_{j=1}^n \omega_{i,j}$ 是数据点 v_i 与所有数据点之间的边的权值之和,如下所示:

$$X(G) = \begin{pmatrix} x_1 & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & x_n \end{pmatrix} \quad (2)$$

加权网络 G 的 Laplacian 矩阵为:

$$L(G) = X(G) - W(G) \quad (3)$$

加权网络 G 的 Laplacian 能量为:

$$E_L(G) = \sum_i \gamma_i^2 \quad (4)$$

其中, $\gamma_1, \gamma_2, \dots, \gamma_i, \dots, \gamma_n$ 表示 $L(G)$ 的特征值。

为计算方便:

$$E_L(G) = \sum_{i=1}^n x_i^2 + 2 \sum_{i < j} \omega_{i,j}^2 \quad (5)$$

任意数据点 v_i 的拉普拉斯中心性为:

$$c_i = \frac{(\Delta E)_i}{E_L(G)} = \frac{E_L(G) - E_L(G_i)}{E_L(G)} \quad (6)$$

其中, $E_L(G_i)$ 为去掉数据点 v_i 后的 Laplacian 能量值。

2.3 最小距离值 δ_i

对于一个数据点 v_i ,用 Laplacian 中心性作为评价其局部重要性的指标值 c_i ,将该数据点的 c_i 值与其他数据点的 Laplacian 中心性数值作比较,得到所有局部重要性比自己高的数据点集合,并计算自己与该集合中每个数据点的距离的最小值 δ_i :

$$\delta_i = \min_{j: c_i > c_j} d_{ij} \quad (7)$$

2.4 确定类簇中心和数据点聚类

计算所有数据点的 Laplacian 中心性 c 和最小距离值 δ ,将它们标示在分别以 c 和 δ 为横坐标和纵坐标的决策图上。坐标系右上角的点表示 c 和 δ 都相对较大的数据点,这些数据点就是类簇中心。如图 2(a)的空手道网络数据集,其决策图为图 1。坐标系右上角的两个点的 c 值和 δ 值都明显比其余节点大,这两点即为类簇中心。

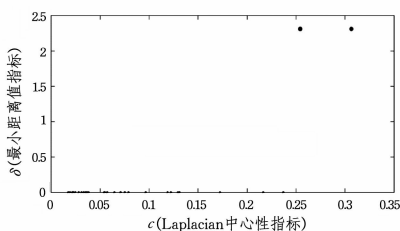


图 1 类簇中心决策图

Fig. 1 Decision graph of cluster center

对所有的类簇中心贴上类簇标签,然后根据未分配类簇

标签的数据点的 c 和 δ 给出其应该归属的类簇。具体方法为:对于某一个没有类簇标签的数据点,将其归属于一个距离最近的有类簇标签的数据点所属的类簇中,比如对于某一数据点 v_i ,计算其 c_j 和 δ_j ,根据 δ_j 找到距离数据点 v_i 最近的局部重要性指标值比 c_j 大的数据点 v_k ,把数据点 v_i 归属到数据点 v_k 所属的类簇,并获得该类簇的标签,循环这个过程直至所有的数据点都被分配到相应的类簇为止,聚类算法完成。

3 数值仿真和结果分析

3.1 数据集

本文将使用以下 6 个数据集来进行数值仿真:1) zachary^[22]空手道俱乐部网络数据集,包含 34 个具有相似性的空手道俱乐部会员,俱乐部分成两个小俱乐部,此处我们把网络中节点之间的路径长度等同于节点之间的距离;2) synthetic_spirals^[24]数据集,包含 250 个数据点和 2 个圆形射线型的类簇;3) synthetic_cassini^[24]数据集,包含 250 个数据点和两个香蕉型类簇(非凸)弯曲在一个圆形的类簇;4) zahn_compound^[24]数据集,包含 399 个数据点和 4 个类;5) fu_flame^[23]模拟火焰数据集包含 240 个数据点和 2 个类簇;6) synthetic_cuboid^[24]三维长方体数据集,包含 250 个数据点和边界位置的 3 个小长方体以及中心位置的小正方体这 4 个类簇。在后面 5 个数据集中,会使用数据点之间的距离作为数据点相互之间的连边的权重值。

3.2 评价指标

本文将使用 Specificity, Rand Index 和 Sensitivity 这 3 个评价指标^[24]来评价聚类效果的好坏,这些指标的的定义的描述如下。

假设一个集合中有 N 个元素,集合中的元素实际有 C 个类别,聚类算法对集合中的元素进行聚类,得到 K 个簇,则 4 个参数分别为:

TP 表示同一个类的元素被分到同一个簇;

FP 表示不同类的元素被分到同一个簇;

TN 表示不同类的元素被分到不同簇;

FN 表示同一类的元素被分到不同簇。

在此基础上,给出如下 3 个聚类指标所示。

1) Specificity 表示不同类元素被正确聚类的比例

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

2) Rand Index 表示正确聚类的比例

$$R = \frac{TP + TN}{TP + FP + FN + TN} \quad (9)$$

3) Sensitivity 估计的是同类元素被正确度量的百分比

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

3 个性能指标都是越接近 1,聚类性能越好。

3.3 结果分析

6 个数据集的聚类结果如图 2 所示。从图 2 可以看出,6 个数据集形状大小和维数各异,如果用经典的 k-means 和 DBSCAN 算法进行实验是无法完成全部数据集的聚类的,而本文的 LDDC 算法却出色地完成了 6 个数据集的聚类,表明了 LDDC 算法的有效性。

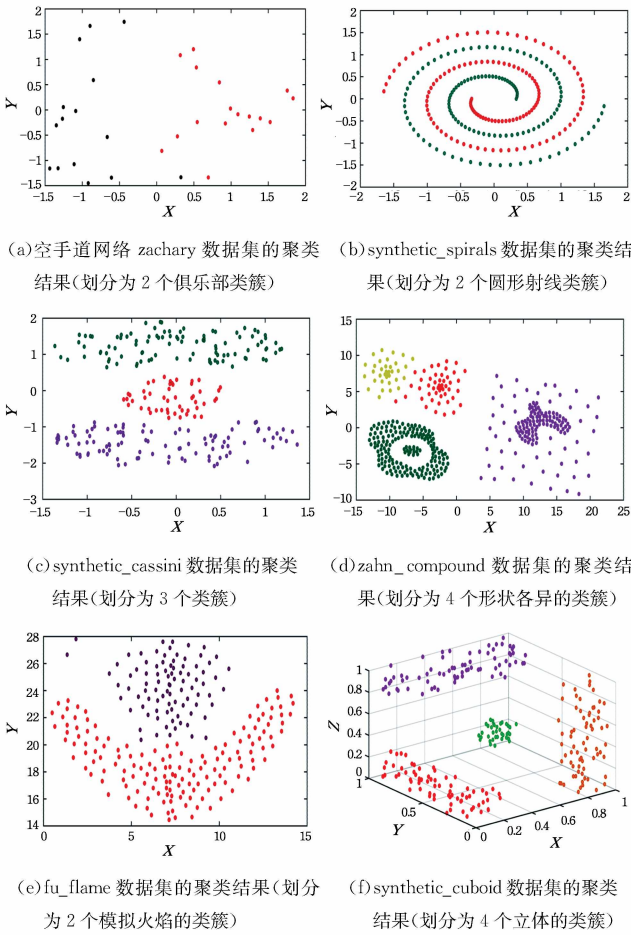


图 2 6 个数据集的聚类结果图
Fig. 2 Clustering result of six datasets

同时,针对这 6 个数据集,用 Specificity, Rand Index 和 Sensitivity 3 个聚类评价指标对 LDDC 与一系列经典聚类算法的性能进行对比,这些算法包括 DDC^[13], k-means^[7], K-medoids^[8], DBSCAN^[10], Affinity^[25], Hierarchical^[26], DensityCut^[27], SpectralClustering^[28], MCODE^[29]。关于这 9 种聚类算法的参数选择问题,我们对算法的参数进行了调试并最终选择了最好的聚类评价结果与本文的算法进行对比分析,实验结果如表 1—表 3 所列。

本文的主要创新点并不是提出一个单纯在聚类效果上超过 DDC 和其他 9 种知名算法的聚类方法,而是想实现一种同时具有如下两个特点的聚类算法:

- 1) 真正无参数,可以在无任何人为设置参数的情况下自动地对数据集进行分类;
- 2) 聚类效果良好,具体的聚类效果与其他经典算法大致等同(或稍有超过)。

DDC 和其他 9 种知名算法的聚类方法是有参数的方法,但目前并无有效的解析方法获得其参数,一般都只是根据经验通过反复调试获得;对于某一算法而言,往往是将所有参数中对应的最好聚类效果作为该算法的聚类结果。而本文提出的 LDDC 算法没有任何人为设置参数的环节,其聚类效果从表 1—表 3 可以看出:在所有的数据集中,LDDC 算法的 3 个性能指标在总体上都高于或者接近 DDC 和其他算法,显示了比较好的聚类效果,实际上就是 LDDC 既在总体上保持了已有 DDC(或其他经典算法)的聚类效果,同时也真正实现了无参数的自动分类。

表 1 不同算法在不同数据集上基于 Specificity 指标的聚类效果对比

Table 1 Comparison of clustering effects on different datasets of different algorithms for Specificity

算法	空手道 zachary	synthetic_spirals	synthetic_cassini	zahn_compound	fu_flame	synthetic_cuboid
LDDC	1.0	1.0	1.0	0.995	1.0	1.0
DDC	1.0	1.0	1.0	0.991	1.0	1.0
k-means	—	1.0	1.0	1.0	1.0	1.0
k-medoids	1.0	1.0	1.0	1.0	1.0	1.0
DBSCAN	1.0	1.0	1.0	1.0	1.0	1.0
Affinity	1.0	1.0	1.0	1.0	1.0	1.0
Hierarchical	1.0	1.0	1.0	1.0	1.0	1.0
DensityCut	—	1.0	1.0	0.935	1.0	1.0
SpectralClustering	0.997	1.0	1.0	0.999	0.998	1.0
MCODE	0.788	0.999	0.999	0.996	0.991	1.0

表 2 不同算法在不同数据集上基于 Rand Index 指标的聚类效果对比

Table 2 Comparison of clustering effects on different datasets of different algorithms for Rand Index

算法	空手道 zachary	synthetic_spirals	synthetic_cassini	zahn_compound	fu_flame	synthetic_cuboid
LDDC	0.94	1.0	1.0	0.96	1.0	1.0
DDC	0.941	1.0	1.0	0.951	0.999	1.0
k-means	—	0.518	0.84	0.889	0.727	1.0
k-medoids	0.9	0.52	0.939	0.889	0.745	1.0
DBSCAN	0.522	1.0	1.0	0.973	0.978	1.0
Affinity	0.695	0.521	0.721	0.819	0.553	0.843
Hierarchical	1.0	1.0	1.0	0.98	0.956	1.0
DensityCut	—	1.0	1.0	0.941	0.999	1.0
SpectralClustering	0.624	0.75	1.0	0.883	0.816	0.843
MCODE	0.56	0.525	0.99	0.902	0.492	0.908

GDB-ICP 方法^[9]使用的是 3.2GHz 的奔腾 4 与 2GB 内存的机器,因此其可比性不大。

LPF-ICP 方法虽然在配准 Rensselaer 数据组上^[9]取得了成功,但仍存在改进的空间。首先,ICP 配准的收敛范围窄,恰当的初始变换往往是其成功的一个关键因素。上述结果也表明,在配准大射影变换的 Capital Region 图对时,由于简单相似变换不是复杂射影变换的一个好的近似,因此需要用更复杂的仿射变换来初始化。一维匹配^[31]就是一个仿射变换的估算方法,上述实验中也初步证实了其有效性。为了给大畸变图像提供适合的初始变换,仍需进一步探讨 LP 种子点与一维匹配方法的对接方法。另外,在实验中,初始引导区的边长一般是定为 200 像素。然而,也观察到两个例外情况:1)在配准重叠区小的 Dashpoint 图对与 Boston Library 图对时,需要将边长相应缩小;2)在配准光照及焦距变换大的 Extreme Zoom 图对与 Summer Night-Winter Day 图对时,为了应对引导区内稀疏的特征点,需加大边长。为了给各类图像的配准确定适合的初始引导区尺寸,需要进一步研究重叠区尺寸、对应特征点数等因素与引导区尺寸的关系等。

结束语 本研究表明,利用 LPCD 方法能从大形貌畸变图像中提取可靠的 LP 角点。进一步,用匹配的 LP 种子点(包括 LP 角点与泡点)而非 SIFT 特征点为图像配准提供可靠的初始引导区。另外,只采用单尺度特征点(Harris 角点及边点)在原始图像上驱动 ICP 过程,就能实现引导区扩张及其配准模型的升级,直至全图像配准。通过整合上面两种技术,本文提出了一种按 LP 种子长大方式来配准二维图像的新方法。实验评价表明,该方法成功地配准了所有 22 对挑战性图像对,而 GDB-ICP 方法则只完成了其中的 19 对,从而证实了 LPF-ICP 方法的有效性。在今后的研究中,将进一步探索提供初始仿射变换的方法与确定初始引导区尺寸的自适应方法,以扩展适用范围;同时也应减少计算开销,以提高实用性。

参 考 文 献

- [1] BROWN M,LOWE D G. Automatic panoramic image stitching using invariant features [J]. *Int. J. Comput. Vis.*, 2007, 71(1): 59-73.
- [2] DUFAUX F,MOSCHENI F. Motion estimation techniques for digital TV: A review and a new contribution[J]. *Proc. of the IEEE*, 1995, 83(6): 858-876.
- [3] LUCAS B D,KANADE T. An iterative image registration technique with an application to stereo vision [C]// *International Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 1981: 674-679.
- [4] BROWN L G. A survey of image registration techniques [J]. *ACM Comput. Surv.*, 1992, 24(4): 325-376.
- [5] BAKER S,MATTHEWS I. Lucas-Kanade 20 years on: A unifying framework [J]. *International Journal of Computer Vision*, 2004, 56(3): 221-255.
- [6] HOMBK P,SCHUNCK B G. Determining optical flow [J]. *Artif. Intell.*, 1981, 17(1-3): 185-203.
- [7] MAES F,COLLIGNON A,VANDERMEULEN D, et al. Multi-modality image registration by maximization of mutual information [J]. *IEEE Trans. Med. Imag.*, 1997, 16(2): 87-198.
- [8] VIOLA P, III W M W. Alignment by maximization of mutual information [J]. *International Journal of Computer Vision*, 1997, 24(2): 137-154.
- [9] YANG G, STEWART C V, SOFKA M, et al. Registration of challenging image pairs: initialization, estimation, and decision [J]. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 2007, 29(11): 1973-1989.
- [10] ZOKAI S,WOLBERG G. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformation [J]. *IEEE Trans. on Image Process.*, 2005, 14(10): 1422-1434.
- [11] LOWE D. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [12] HARRIS C,STEPHENS M. A combined corner and edge detector [C]// *Proc. of 4th Alvey Vision Conference*. Manchester, United Kingdom: 1988: 147-151.
- [13] CANNY J. A computational approach to edge detection [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679-698.
- [14] FAUGERAS O,LUONG Q T,PAPADOPOULOU T. The geometry of multiple images [M]. Cambridge, MA: MIT Press, 2001: 251-255.
- [15] CAN A, STEWART C, ROYSAM B, et al. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina [J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, 24(3): 347-364.
- [16] STEWART C, TSAI C L, ROYSAM B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration [J]. *IEEE Trans. on Medical Imaging*, 2003, 22(11): 1379-1394.
- [17] PRESS W H, TEUKOLSKY S A, VETTERLING W T, et al. *Numerical recipes in C: The art of scientific computing* (2nd ed) [M]. Cambridge CB2 1RP: Cambridge University Press, 1992: 683-688.
- [18] TSAI C L, LI C Y, YANG G, et al. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multi-modal fluorescein angiogram sequence [J]. *IEEE Transactions on Medical Imaging*, 2010, 29(3): 636-649.
- [19] YANG S, YANG C L. Image registration algorithm based on sparse random projection and scale-invariant feature transform [J]. *Acta Optica Sinica*, 2014, 34(11): 98-102. (in Chinese)
杨飒, 杨春玲. 基于压缩感知与尺度不变特征变换的图像配准算法 [J]. *光学学报*, 2014, 34(11): 98-102.
- [20] XU J J, ZHANG Y, ZHANG H. Fast image registration algorithm based on improved Harris-SIFT descriptor [J]. *Journal of Electronic Measurement and Instrumentation*, 2015, 29(1): 48-54. (in Chinese)
许佳佳, 张叶, 张赫. 基于改进 Harris-SIFT 算子的快速图像配

- 准算法[J]. 电子测量与仪器学报, 2015, 29(1): 48-54.
- [21] BESL P, MCKAY N. A method for registration of 3 D shapes [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1992, 14(2): 239-256.
- [22] ZOU X J, LIN G C, TANG Y C, et al. Sub-Pixel object-Image Registration Using Improved Iterative Closest Point Method [J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(8): 1242-1249. (in Chinese)
邹湘军, 林桂潮, 唐昶超, 等. 改进迭代最近点法的亚像素级零件图像配准[J]. 计算机辅助设计与图形学学报, 2016, 28(8): 1242-1249.
- [23] DAI J L, CHEN Z Y, YE X Z. The Application of ICP Algorithm in Point Cloud Alignment [J]. Journal of Image and Graphics, 2007, 12(3): 517-521. (in Chinese)
戴静兰, 陈志杨, 叶修梓. ICP算法在点云配准中的应用[J]. 中国图象图形学报, 2007, 12(3): 517-521.
- [24] LONG J W, SHEN X J, ZANG H, et al. An Adaptive Thresholding Algorithm by Background Estimation in Gaussian Scale Space [J]. Acta Automatica Sinica, 2014, 40(8): 1773-1782. (in Chinese)
龙建武, 申铨京, 臧慧, 等. 高斯尺度空间下估计背景的自适应阈值分割算法[J]. Acta Automatica Sinica, 2014, 40(8): 1773-1782.
- [25] LIU L, PENG F Y, ZHAO K, et al. Simplified SIFT algorithm for fast image matching [J]. Infrared and Laser Engineering, 2008, 37(1): 181-184. (in Chinese)
刘立, 彭复员, 赵坤, 等. 采用简化 SIFT 算法实现快速图像匹配 [J]. 红外与激光工程, 2008, 37(1): 181-184.
- [26] TAO T, ZHANG Y. Detection and description of scale-invariant keypoint in log-polar space [J]. Journal of Image and Graphics, 2015, 20(12): 1639-1651. (in Chinese)
陶涛, 张云. 对数极坐标系下尺度不变特征点的检测与描述 [J]. 中国图象图形学报, 2015, 20(12): 1639-1651.
- [27] MIKOLAJCZYK K, SCHMID C. Scale & affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, 60(1): 63-86.
- [28] HOLLAND P W, WELSCH R E. Robust regression using iteratively reweighted least-squares [J]. Commun. in Statist. -Theor. and Meth., 1977, 6(9): 813-827.
- [29] BURNHAM K P, ANDERSON D R. Model selection and inference: A practical information-theoretic approach [M]. Berlin, Germany: Springer, 1998.
- [30] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision [M]. Cambridge, UK: Cambridge Univ. Press, 2003: 148-149.
- [31] ZHANG Y, CHU C H. One-dimensional mapping for estimating projective transformations [J]. IEEE Trans. on Image Process, 2010, 19(11): 3049-3058.

(上接第 296 页)

- [19] ZHAO X W, LIANG J Y. A hybrid data attribute weighted clustering algorithm based on the information entropy [J]. Journal of Computer Research and Development, 2016, 53(5): 1018-1028. (in Chinese)
赵兴旺, 梁吉业. 一种基于信息熵的混合数据属性加权聚类算法[J]. 计算机研究与发展, 2016, 53(5): 1018-1028.
- [20] QI X, FULLER E, LUO R, et al. A novel centrality method for weighted networks based on the Kirchhoff polynomial [J]. Pattern Recognition Letters, 2015, 58(C): 51-60.
- [21] QI X, FULLER E, WU Q, et al. Laplacian centrality: A new centrality measure for weighted networks [J]. Information Sciences, 2012, 194(5): 240-253.
- [22] ZACHARY W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropological Research, 1977, 33(4): 452-473.
- [23] FU L, MEDICO E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data [J]. BMC Bioinformatics, 2007, 8(1): 3.
- [24] WIWIE C, BAUMBACH J, RÖTTGER R. Comparing the performance of biomedical clustering methods [J]. Nature Methods, 2015, 12(11): 1033.
- [25] HUANG D C, QIAN C K. Based on dimension attribute distance mixed attribute affinity propagation clustering algorithm [J]. Computer Science, 2015, 42(Z11): 55-57. (in Chinese)
黄德才, 钱潮恺. 基于维度属性距离的混合属性近邻传播聚类算法 [J]. 计算机学报, 2015, 42(Z11): 55-57.
- [26] NAVARRO J, FRENK C, WHITE S M. A Universal Density Profile from Hierarchical Clustering [J]. Astrophysical Journal, 2009, 490(2): 493-508.
- [27] DING J, SHAH S, CONDON A. densityCut: an efficient and versatile topological approach for automatic clustering of biological data [J]. Bioinformatics, 2016, 32(17): 2567-2576.
- [28] WANG Y L, CHEN B, WANG X L, et al. Improved adaptive spectral clustering algorithm based on density adjustment [J]. Control and Decision, 2014(9): 1683-1687. (in Chinese)
王雅琳, 陈斌, 王晓丽, 等. 基于密度调整的改进自适应谱聚类算法 [J]. 控制与决策, 2014(9): 1683-1687.
- [29] BADER G D, HOGUE C W. An automated method for finding molecular complexes in large protein interaction networks [J]. BMC Bioinformatics, 2003, 4(1): 2.