2010年 7月 第 7期
Vol. 31 No. 7 2010

多元图表示原理的散点图分类器设计

张 涛[1,2], 洪文学[2]

[1](燕山大学 信息科学与工程学院, 河北 秦皇岛 066004)
[2](燕山大学 电气工程学院, 河北 秦皇岛 066004)
E-mail: zhtao_79@163.com

摘 要: 基于多元图表示原理, 提出散点图可视化分类器. 该分类器的基本思想是将数据矩阵映射为散点图, 利用散点图表示的紧致性, 通过像素图将其转换为图像并利用图像扩展形成分类子空间, 最终将多个分类子空间按照组合分类规则构成组合分类器. 该分类器集成了图表示技术与图像处理技术, 使整个分类过程可视, 可实现交互式分类. 利用 Iris和 Wine数据集的实验表明, 散点图分类器对训练样本数量敏感度较低, 分类性能接近甚至优于目前的主流分类器.

关 键 词: 分类器; 散点图; 图表示; 图像处理; 交互式
中图分类号: TP391 文献标识码: A 文 章 编 号: 1000-1220(2010)07-1433-06

Scatter Classifier Design Based on Multivariate Graphical Representation Theory

ZHANG Tao[1,2], HONG Wen-xue[2]

[1](College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)
[2](College of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China)

Abstract: A novel visual combining classifier based on graphical representation named scatter classifier has been proposed. The basic principle of that is as follows: at first mapping data matrix to graph by scatter plot and then using pixel graph convert the graphs space into images space, the sub classifiers are obtained from the images by pixel expanding. At last different sub classifiers are combined as a combining classifier by combining rule. The classifying process of scatter classifier is visual and interactive, which is the result that the classifier takes advantages of visualization of graphical representation and introduces the image processing technology. A serious of experiments based on Iris dataset and Wine dataset has been done and the experiments show that scatter classifier is insensitive for the size of training set, and the performance of that is close to or even outperformed many popular classifiers.

Key words: classifier; scatter; graphical representation; image processing; interactive

1 引 言

近年来, 可视化技术在模式识别领域中得到突破式的增长[1,2]. 可视化技术的最初目的为信息可视化, 即通过图形或图像的方式表现抽象的数据, 从而使数据更容易被人进行研究和处理. 由于模式识别领域中处理的数据一般为多维数据, 即多元统计分析中的多元数据, 因此以散点图、雷达图、平行坐标等为代表的多元图表示(简称图表示)方法成为了数据分析的主要可视化方法[3].

随着可视化分析学(visual analytics)[4]概念的提出, 可视化的任务也由对数据本身的关注, 开始向数据分析与知识获取过渡. P. Compieta指出[5]: 将可视化融入数据挖掘与知识发现, 可以增强系统与用户的交互性与过程的可解释性. 传统的多元可视化以数据为核心, 关注多元数据空间到一维、二维或者三维图像视觉空间的匹配, 并且最大程度保留原始数据的结构信息, 符合模式识别领域中对数据表示的要求, 因此在模式识别领域获得了新的发展[6]. 以多元图表示为基础的模式识别理论, 在数据降维、信息融合、特征选择等方面已经得到了比较深入的研究[7-9], 但基于多元图表示的分类器设计还处于起步阶段[6].

目前基于图表示原理的分类器主要分为两类: 一类是利用图表示方法对原始数据进行特征选择或特征提取, 再利用传统的分类器进行分类, 如文献[10]中设计的基于雷达图特征提取的分类器. 该类分类器采用的图表示方法一般为多点表示, 如雷达图、脸谱图等, 利用图表示对数据的图形化描述, 从图形中提取视觉特征作为数据表示. 但在分类阶段仍采用 kNN等传统的分类方法, 可视化过程仅局限于数据的表示阶段, 并非真正意义上的图表示分类器.

另一类分类器是从数据的表示到分类过程完全基于图表示方法, 具有很好的可视化的特点. 但该类方法目前主要集中于以平行坐标作为图表示方法平行坐标分类器[11]. 平行坐标虽然是信息可视化的主流技术之一, 但对一般用户而言不易理解, 从而影响交互的效果. 而且平行坐标属于多点表示的图表示方法, 随着数据的增加, 图形会变得凌乱, 不利于有用信

收稿日期: 2009-04-07 收修改稿日期: 2009-06-16 基金项目: 国家自然科学基金项目(60671025, 60873121, 60904100)资助. 作者简介: 张 涛, 男, 1979年生, 博士研究生, 讲师, 研究方向为模式识别, 图像处理; 洪文学, 男, 1953年生, 工学硕士, 教授, 博士生导师, 研究方向为可视化模式识别, 信息融合.

息的发现[9].

针对以上两种分类器的不足,本文设计一种新的图表示分类器—散点图分类器.该分类器以散点图作为数据分类的基础.与其他图表示方法相比,散点图简单直观,更容易被用户理解和接受,因此使用广泛.更重要的是,散点图为数据的单点表示,不会随着表示的样本数增加而凌乱,且符合Aankerst提出的分类器表示基本原则[12],这为分类器的设计奠定了基础.另外,本文在分类界面的计算上引入图像处理技术,使分类过程符合人类的视觉分类习惯,充分发挥散点图在表示阶段的可视化特色.因此,散点图分类器从数据的表示到分类过程符合人类的视觉习惯,实现了全过程的可交互性与可解释性,有利于人类知识与模式识别系统的集成和新知识的发现,满足可视化分析学对分类器的要求.

2 散点图分类器基本原理

散点图分类器的基本原理如图1所示.由图可以看出,该分类器由3部分组成:散点图表示,像素图表示和分类器组合.散点图表示负责将多维数据矩阵的列向量(训练样本的同一属性)映射为散点图表示,将抽象的数据转化为直观的图.像素图表示则将散点图转换为像素图并利用扩展算法进

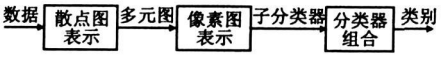


图1 散点图分类器框图

Fig. 1 Diagram of the scatter classifier

行连通区域扩展,利用混合色表示类别的重叠情况.扩展后的图像可以视为当前数据集的一个分类空间或子分类器.分类器组合则是将这些子分类器集成为一个组合分类器,实现最终的分类.由于整个分类过程均可以进行可视化表示,因此该分类器属于可视化组合分类器.

2.1 多维数据的预处理

设矩阵 X 表示多维数据集合

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \ddots & \cdots & \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

其中,x_{ij} 为第 i 个样本的第 j 个属性.$[x_{i1}, x_{i2}, x_{i3}, \cdots x_{im}]$ 为同一样本的 m 个属性.$[x_{1j}, x_{2j}, x_{3j}, \cdots x_{nj}]^{T}$ 是所有样本第 j 个属性的集合,$[\cdot]^{T}$ 为向量的转置.为了将数据在特定的区域内进行图表示,简化后期的处理复杂度,因此在预处理阶段需对矩阵 X 作归一化处理.

在分类过程中,本文采用列向量(即不同数据样本的同一属性)对数据进行分析.这样的分析方法可以克服行向量间的不可比性与不确定性,符合相对描述与概念描述的学术思想[13].因此采用极差归一化方法,如式(1)所示:

$$x^{*}_{ij}=\frac{x_{ij}-\min\limits_{1\leqslant i\leqslant n} x_{ij}}{R_j} \tag{1}$$

其中,$R_j=\max\limits_{i=1,\cdots,n} x_{ij}-\min\limits_{i=1,\cdots,n} x_{ij}$ 表示极差,归一化后的数据分布范围是 $[0, 1]$.

2.2 数据的散点图表示与优化

散点图是数据可视化中最常用的方法之一,它可以显示两个变量之间的关系.在散点图中,每个数据样本对应一个点或者标记,其位置坐标由两个变量的值决定.通过散点图可以观察和理解聚类,离群点,趋势以及相关等数据结构信息.二维数据的直角散点图实际上就是以二维数据变量为坐标在平面直角坐标系中描点表示,是散点图中最简单和容易理解的一种,如没有特殊声明,本文所指散点图均为直角散点图.对于完成列归一化的数据矩阵 $[x^{*}_{ij}]$,其对应的散点图坐标为

$$\begin{cases} x=x^{*}_{ij} \\ y=x^{*}_{ik} \end{cases} \tag{2}$$

以Iris数据集的 sepal length与 sepal width特征结合为例,其未经优化的散点图表示如图2(a)所示.由该图可以看出,散点图清晰的表示出了该数据集不同类别间的空间分布关系,符合用于分类的表示基本原则[10],为后期的分类过程打下了良好的基础.

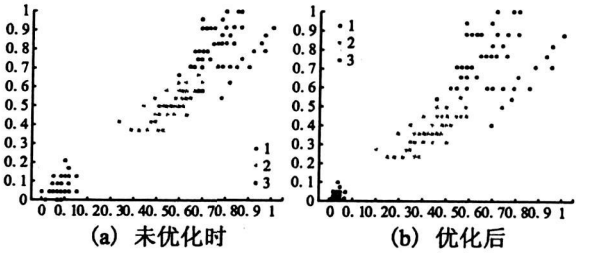


图2 Iris数据集 sepal length与 sepal width特征散点图的优化前后对比

Fig. 2 Compare of original scatter and optimized one for sepal length and sepal width of Iris

同时,由于散点图表示的是特征间的相互关系,因此得到的散点图个数 p 要高于原始的数据维数 m,理论上将达到 m^2,考虑到数据的对称性,对其约简之后仍达到 $\binom{m}{2}$,相当于对特征向量进行了升维表示.为了防止"组合爆炸(combinatorial explosion)"问题的产生,可以通过设置使用散点图个数缺省值来对多元图数目进行约束.

散点图表示虽然可以直观的表示数据的分布,但该分布不一定适合分类,因此需要进行非线性优化.非线性函数 $f(x)$ 的选取原则为:

条件1:$f(x)\in[x_{min}, x_{max}]$,当 $x\in[x_{min}, x_{max}]$

条件2:如果 $x_1\geqslant x_2$,则有 $f(x_1)\geqslant f(x_2)$ 其中 $x_1\in[x_{min}, x_{max}]$ 且 $x_2\in[x_{min}, x_{max}]$

条件1限制了 $f(x)$ 的值域范围,要求与定义域相同,其目的是保证经过非线性变换之后数据分布区间不发生变化,否则可能需要对后续过程重新调整.条件2要求 $f(x)$ 为单调函数,以避免数据相对关系的混乱,影响表示的紧致性.理论

上, 任何满足该条件的函数均可用于散点图的非线性优化.

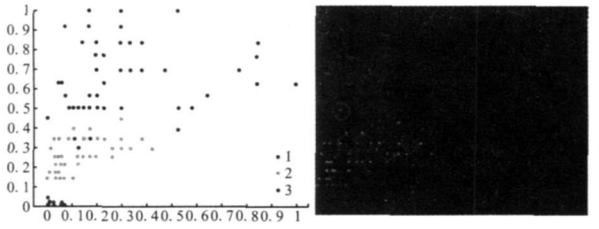
本文利用多项式函数 $f(x)=x^a$ 完成该优化过程. 优化后的坐标表达为

$$\begin{cases} x=x_{ij}^{a_1} \\ y=x_{ik}^{a_2} \end{cases} \tag{3}$$

式中, a_1, a_2分别表示对原始数据的非线性参数. 要求 $a_1>0$ 且 $a_2>0$以满足条件 2的要求, 同时排除 $a=$的情况以避免数据同一化. 对图 2(a)的非线性优化结果如图 2(b)所示. 通过对比可以看到, 经过非线性优化之后, 完全可分的 setosa类数据(红色点表示)更为集中的会聚到原点附近, 而出现一定混叠区域的 versicolor类(绿色点表示)与 virginica类数据(蓝色点表示)边界处数据点间间距变大, 这有利于分类界面的确定. 因此, 经过非线性优化后的散点图表示无论从直观识别上还是从理论分析上, 都更有利于分类.

2.3 像素图与区域扩展

散点图的最初目的是用于数据的表示, 而不是用于数据的分类, 而且传统的散点图表示方法无法从空间分布和类别概率两个角度同时对数据进行直观表示. 本文结合散点图表示与概率分布模型, 提出基于色度学的类别分布像素图(简称像素图)的概念, 用于深化散点图表示中不同类别数据的概率关系, 使散点图表示可以直观的描述类别分布情况, 为散点图在可视化分类中的应用提供良好的工具.



(a) 散点图　　(b) 像素图

图 3　散点图与像素图比较

Fig. 3　Compare of a scatter and its pixel graph

在色度学中, 色度空间的任何颜色均可由基色按一定比例合成. 一般情况下, 选择的基色为红色、绿色和蓝色. 在像素图中, 基色表示相应的类别, 而重合点则由该点按类别概率根据配色原理使用混合色表示. 当表示的类别超过 3个时, 可以利用补色作为新的基色, 并利用新的混合色与饱和度表示. 图 3表示了散点图及和它的像素图. 为了清楚的表示出重叠点, 将混合色点利用红色圆圈进行了标记. 通过比较, 可以发现像素图不但直观的表示的样本的空间分布, 更重要的是, 直观的表示了重合点的类别分布情况. 因此, 像素图更适合于分类的情况. 更为重要的是, 像素图更有利于系统的可视化, 从而带来更好的交互式效果.

通过像素图, 可以更好的对数据进行可视化, 但分类的最终目标是要获得分类界面. 传统模式识别方法中, 一般通过数学方法或几何方法求解分界面. 这种方法虽然具有良好的理论基础, 但未考虑人类视觉因素. 如果采用此方法完成散点图分类器分类界面的求取, 将削弱前期的可视化特色. 因此, 本文初步设计了基于图像扩展原理的分类器分类界面计算, 其规则如下:

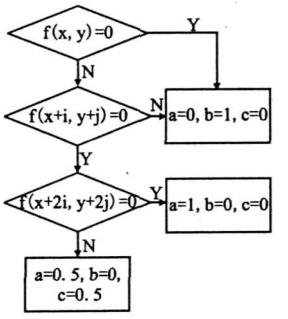


图 4　权重确定流程图

Fig. 4　Flow chart for weighting

将像素图视作一幅图像, 用 $f(x, y)$表示坐标为(x, y)处像素的色度值, 并用 $f(x, y)=0$表示坐标为(x, y)处像素的色调为黑色. 对于 $f(x, y)$的邻域像素 $f(x+i, y+j)$, 区域扩展后该像素点色度可表示为

$$g(x+i, y+j)=af(x, y)+bf(x+i, y+j)+cf(x+2i, y+2j) \tag{4}$$

其中, 加权系数 a, b, c的大小可根据图 4的流程进行确定. 对该过程进行循环, 直至图像中所有的点均非零(即图像中不再存在黑色点, 对应于类空间中所有区域均对应类别及相应概率), 即获得数据空间的模式分类结构, 分类器的设计也由图形空间过渡至类空间. 对图 3(b)扩展后的图像如图 5所示, 其可以认为是一个子分类器. 对于任意未知类别的样本,

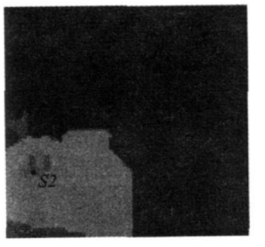


图 5　子分类器示例

Fig. 5　An example of sub classifier

仅通过对比其散点图坐标在对应分类器上的颜色即可判断其所属类别及概率. 如对于未知样本 s_1 由于其对应颜色为单纯的蓝色, 因此可判断为 virginica类(蓝色表示); 而对于未知样本 s_2 通过颜色可以直观的判断其为蓝色与绿色的混合, 且更偏向于绿色, 因此其属于 versicolor类(绿色表示)的可能性更大一些. 如果读取 s_2点对应的颜色数据, 可以发现其对应的颜色为青色偏绿, 属于绿色与蓝色按 2∶1的比例混合得到, 因此其属于 versicolor类的概率为 67.7%, 而属于 virginica类的概率为 33.3%. 通过该分类过程可以看到, 散点图分类器既适用于直观的通过观察方式的分类方法, 也适合

于传统的利用概率密度方式的分类方法.

2.4 分类器组合

区域扩展后的像素图实际上已经从特征空间过渡到了类空间, 所获得的分类模式具有良好的可视化特色, 可以直接根据测试样本坐标所对应的颜色值获得其所属类别及概率信息. 因此, 基于区域扩展技术的散点图分类器是一种基于多元图表示的可视化分类器, 可以作为分类器直接使用. 同时, 对于特征向量维数高于 1维的数据, 其列向量(即样本的属性)散点图表示必然可以由多个散点图共同完成表示(比如 Iris数据集可以由 6个散点图表示), 可以认为是由多个分类器共同完成分类任务的组合分类器. 因此, 散点图分类器是典型的可视化组合分类器, 其实现过程如图 6所示.

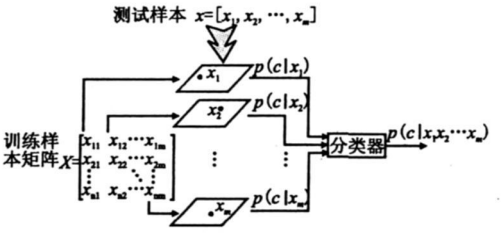


图 6 散点图分类器分类过程

Fig. 6 Realized scheme of scatter classifier

对于待分类样本, 依训练样本参数进行预处理与散点图坐标计算, 并与对应的子分类空间进行比较, 获得该样本属于 c类的概率 $P(c|x_i)$, 最后通过判决规则对各子分类器的判决结果进行分析, 获得最终判决结果. 在本文实验中, 组合规则采用的是和式规则:

$$P(c|x_1 x_2 \cdots x_m)=\sum_{i}^{m}\alpha_i P(c|x_i) \qquad (5)$$

其中, α_i表示第 i个子分类器的加权系数, 由混合色与总像素数比例获得, 因此 $\alpha \in [0, 1]$.

3 实验与分析

在本节中, 将对散点图分类器性能进行一系列的测试, 以验证其有效性.

3.1 测试数据集

本文选择的测试数据集为 Iris数据集和 Wine数据集, 这两个数据集来源于应用广泛的标准模式识别 UCI数据库, 是评价分类算法优劣的通用数据.

表 1 数据集相关信息

Table 1 Some information about datasets

数据集	类别数	属性数	样本数	散点图数目	参考文献
Iris	3	4	150	6	[10, 14, 15, 16, 17]
Wine	3	13	178	78	[14, 15, 16, 17]

其中, Iris数据分为 3类, 共有 150个样本, 每类各包含 50个样本, 样本数量分布均衡, 每个样本有 4个属性. 而 Wine数据集为意大利三种葡萄酒的化学分析数据, 每个样本有 13个属性用于分类, 整个数据集样本可以分为 3类, 对应的样本数分别为 59, 71和 48, 共 178个样本. 与 Iris数据相比, 其属性数目升高, 且各类别样本数不均衡, 因此与 Iris数据结合, 可以对分类器性能做出更客观的评测. 表 1给出了两个数据集的对比和 3.3节中部分实验数据来源.

3.2 散点图数目选择

一般情况下, 每幅散点图表示两个属性之间的关系, 因此可能产生组合爆炸问题. 为了降低组合爆炸带来的计算复杂度和分类复杂度, 可以通过对散点图数目进行设置, 即对生成的散点图进筛选. 筛选的方法可以分为两类, 一类是利用散点图分类器良好的可视化特性, 根据生成的散点图子分类空间进行交互式筛选, 由用户决定最后参与分类的子分类器. 该方法可以很好的集成人类的分类能力, 但参与交互的用户不同, 可能导致最终的分类结果不同, 因此本文实验不采用这种方法; 另一类方法是根据子分类空间中混合色所占比例进行自动加权与筛选, 利用该方法进行分类得到的分类精度可能不如交互式分类得到的精度高, 但结果仅与数据本身有关, 因此具有客观性和可重复性的特点. 为了保证实验结果的可信度, 本文实验采用的筛选与加权方法均为自动加权与筛选.

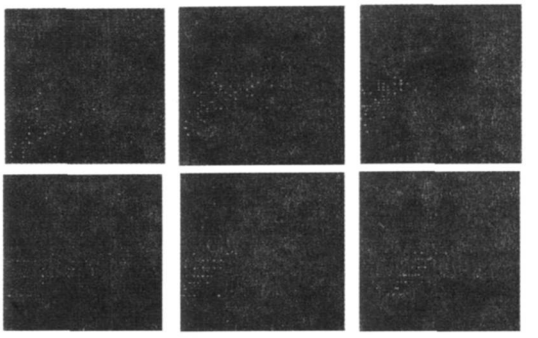


图 7 像素图表示

Fig. 7 Pixel representation

对于 Iris数据集, 由于其每个样本具有 4个属性, 因此其可用的散点图个数为 $\binom{4}{2}=6$幅, 其像素图与对应的分类空间分别如图 7与图 8所示. 利用留一法对不同散点图数目下的

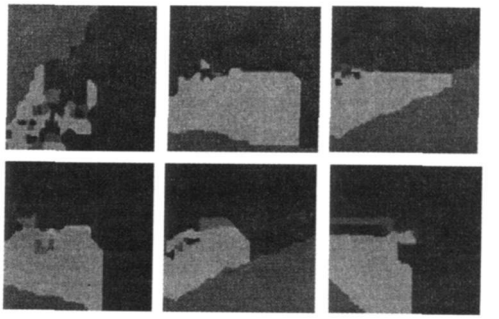


图 8 子分类器空间

Fig. 8 Sub classifier space

散点图分类器进行测试, 得到的分类精度如图 9所示. 可以看出, 对于 Iris数据, 不同的散点图数目下分类精度的分布区间

为 96.00% ~98.67%, 不同的散点图数目对于分类精度影响有限.

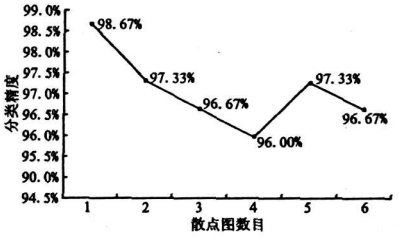


图 9　Iris数据不同散点图数目下散点图分类器的分类精度

Fig. 9　Precision to Iris dataset of scatter classifier under different number of scatters

对于 Wine数据集, 每个样本有 13个属性, 因此其散点图个数达到 $\binom{13}{2}=78$个. 在不同的散点图数目下, 分类器的分类精度如图 10所示, 分类精度的分布区间为 90.45% ~98.31%. 通过 Iris与 Wine的对比实验可以看到, 虽然散点图数目对于分类器的分类精度影响不同, 但都保持了较高的分类精度, 而分类的计算复杂度会随着散点图数目的增加而增加. 因此具体采用的散点图数目需要根据不同的数据集与应用场合来确定.

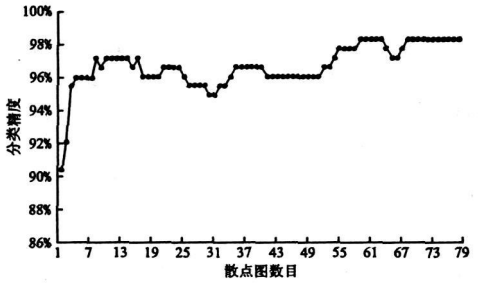


图 10　Wine数据不同散点图数目下散点图分类器的分类精度

Fig. 10　Precision to Wine dataset of scatter classifier under different number of scatters

3.3　交叉验证实验

在实验方法上, 采用 N倍交叉验证 (N-fold cross validation)与留一法交叉验证 (leave one out cross validation, LOOCV)结合进行测试. 其中, N倍交叉验证为将全部数据样本分为 N份, 轮流将其中 N-1份做训练 1份做测试, N次的结果的均值作为对算法精度的估计. N倍交叉验证可以通过对 N的调整分析不同训练样本情况下分类器的分类性能. 由于 N倍交叉验证的结果会因数据集的划分不同而不同, 因此本文实验采用多次交叉验证求均值, 例如 10次 10倍交叉验证, 以保证结果的可信度.

同时, 为了确保分类性能公平比较, 并避免训练集和测试集的依赖, 分类器精度的估计引入留一法交叉验证. 留一法是指数据集共有 M个样本, 使用 (M-1)个样本设计分类器, 并估计剩余的一个样本, 数据集需要重复 M次. 这种估计虽然计算量大, 但是无偏的. 该方法被认为是精度测评的基本方法之一, 且比交叉验证更可靠[18], 被广泛采用.

作为对比算法, 本文选取了多种分类器用作性能的对比测试. 包括经典的 LDA, kNN和 SVM分类器, 基于图表示的雷达图分类器和两种模糊分类器. 其中, LDA为典型的线性分类器, 而 kNN为典型的非线性分类器, 它们与散点图分类器类似, 具有良好的解释性; 而 SVM则具有良好的学习能力, 是目前研究最多的分类器之一; 文献[10]中使用的中使用雷达图分类器是目前性能较好的图表示分类器; 而文献[16]、[17]提供了两种新的模糊分类器, 文献[16]的模糊分类器是基于规则的分类, 具有良好的解释性和简洁性; 文献[17]则是利用了遗传算法实现了模糊分类器最优化的参数提取.

表 2给出了不同测试方法下散点图分类器精度. 在该表中, 由于部分文献仅使用了留一法作为精度估计方法, 因此部分内容为空. 而不同的文献对精度的保留位数不同, 因此出现了小数点位数不同的情况.

通过该实验可以看到, 在不同的训练样本数目下, 散点图分类器的分类精度受到的影响有限, 未出现明显的 "欠学习 "与 "过学习 "现象. 这主要是由于散点图分类器的分类过程与人类分类新样本的过程类似, 具有良好的可解释性, 从而获得了良好的泛化能力, 这也是可视化分类的重要特点之一.

表 2　交叉验证对比实验结果

Table 2　Results under cross validation

分类器	Iris数据集			Wine数据集		
	5倍交叉验证	10倍交叉验证	留一法	5倍交叉验证	10倍交叉验证	留一法
散点图分类器	98.00%	98.00%	98.67%	97.19%	96.67%	98.31%
LDA	98.00%	98.00%	98.00%	97.75%	98.88%	98.88%
kNN, Euclidean, k=1	95.33%	96.00%	96.00%	94.44%	95.5%	96.1%
Standard LS-SVM	94.66%	96.59%				
雷达图分类器	-	-	97.33%	-	-	-
文献[16]方法	-	-	96.00%	-	-	96.07%
文献[17]方法	-	-	95.5%	-	-	99.4%

同时, 散点图分类器的分类精度已经接近甚至超过了经典的 kNN和模糊分类器, 证明了其分类性能的可用性.

4　结　论

基于散点图的可视化和易于理解的特点, 本文设计了一

种新的可视化组合分类器-散点图分类器.在分类过程中,散点图简单直观的表示特性分类易于理解的基础,而数据表示的紧致性为分类奠定了理论基础;基于色度学的类别分布像素图概念的引入强化了散点图的分类功能,而图像处理技术的使用保证了整个分类过程的可视化与用户友好;最终,权重的计算与子分类器的组合解决了多个散点图可能带来的组合爆炸问题,降低了分类的复杂度.因此,散点图分类器是集成多元图表示、图像处理和模式识别技术于一体的可视化组合分类器,具有易于理解、可交互式分类的特点.

为了验证散点图分类器的有效性,本文利用UCI数据集中的Iris与Wine数据集进行了实验.实验结果表明该分类器具有良好的泛化能力和较高的分类精度,证明了其可行性.同时,实验也表明该分类器性能仍具有一定的提升空间,如何在不影响其可视化特色的基础上提高分类性能,是下一步要深入研究的问题.

References:

[1] Janez Demsar, Gregor Leban, Blaz Zupan. FreeViz-An intelligent multivariate visualization approach to explorative analysis of biomedical data[J]. Journal of Biomedical Informatics, 2007, 40(6): 661-671.

[2] Joshua New, Wesley Kendall, Jian Huang, et al. Dynamic visualization of gene coexpression in systems genetics data[J]. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(5): 1081-1094.

[3] Peter C C Wang. Graphical representation of multivariate data[M]. London: Academic Press, 1978.

[4] Daniel A Keim, George G Robertson, Jim J Thomas, et al. Special section on visual analytics[J]. IEEE Transactions on Visualization and Computer Graphics, 2006, 12(6): 1361-1362.

[5] Compieta P, DiMartino S, Bertolotto M, et al. Exploratory spatio-temporal data mining and visualization[J]. Journal of Visual Languages & Computing, 2007, 18(3): 255-279.

[6] Hong Wen-xue, Li Xin, Xu Yong-hong, et al. Information fusion and pattern recognition based on graphical representation theory[M]. Beijing: National Defence Industry Press, 2008.

[7] Liu Wen-yuan, Meng Hui, Hong Wen-xue, et al. A new method for dimensionality reduction based on multivariate feature fusion[C]. IEEE ICIT, 2007: 108-111.

[8] Xu Yong-hong, Hong Wen-xue, et al. Visual pattern recognition method based on optimized parallel coordinates[C]. IEEE ICIT, 2007: 127-132.

[9] Xu Yong-hong, Hong Wen-xue, et al. Parallel dual visualization of multidimensional multivariate data[C]. IEEE ICIT, 2007: 263-268.

[10] Liu Wen-yuan, Li Fang, Hong Wen-xue. Research on classifier of multi-dimensional data based on radar chart mapping[J]. Computer Engineering and Applications, 2007, 43(22): 161-164.

[11] Harri Siirtola, Kari-Jouko Raiha. Interacting with parallel coordinates[J]. Interacting with Computers, 2006, 18(6): 1278-1309.

[12] Arkedev A G, Braverman E M. Computers and pattern recognition[M]. Washington DC: Thompson, 1966.

[13] Elżbieta Pękalska, Robert P W. Duin, Pavel Paclik. Prototype selection for dissimilarity-based classifiers[J]. Pattern Recognition, 2006, 39: 189-208.

[14] Raghuraj R, Lakshminarayanan S. Variable predictive models-a new multivariate classification approach for pattern recognition applications[J]. Pattern Recognition, 2008, doi: 10.1016/j.patcog.2008.7.5.

[15] Emre Comak, Ahmet Arslan. A new training method for support vector machines: Clustering k-NN support vector machines[J]. Expert Systems with Applications, 2008, 35: 564-568.

[16] Li Jie, Deng Yi-ming, Shen Shi-tuan. Classification rule extraction based on fuzzy area distribution and classification reasoning algorithm[J]. Chinese Journal of Computers, 2008, 31(6): 934-941.

[17] Enwang Zhou, Alireza Khotanzad. Fuzzy classifier design using genetic algorithms[J]. Pattern Recognition, 2007, 40(12): 3401-3414.

[18] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]. Proc of the 14th Int Joint Conference on Artificial Intelligence, 1995: 1137-1143.

附中文参考文献:

[6] 洪文学,李 昕,徐永红,等.基于多元统计图表示原理的信息融合和模式识别技术[M].北京:国防工业出版社,2008.

[10] 刘文远,李 芳,洪文学.基于多维数据雷达图表示的图形分类器研究[J].计算机工程与应用,2007,43(22):161-164.

[16] 李 洁,邓一鸣,沈士团.基于模糊区域分布的分类规则提取及推理算法[J].计算机学报,2008,31(6):934-941.