

二次映射和遗传算法用于鉴别可视化特征提取

王金甲^{1,2}, 李 静^{2,3}, 张 涛^{1,2}, 洪文学²

(1.燕山大学信息学院, 秦皇岛 066004; 2.燕山大学 电气学院, 秦皇岛 066004; 3.燕山大学 理学院, 秦皇岛 066004)



摘 要: 多元数据图表示是高维数据可视化最简单的一种处理方法。从多元数据的雷达图中提出了一种图形特征—可视化重心特征。由于雷达图与数据的特征排序有关, 导致可视化特征深受特征排序影响, 提出了利用二次映射计算出所有特征排序下的可视化特征, 基于遗传算法再从中选择出具有鉴别能力的可视化特征。葡萄酒、乳腺癌和糖尿病等 UCI 真实数据集的实验结果证实了我们的想法, 最佳分类错误率分别达到了 0%、1.61%和 20.7%, 优于报道的常用的分类性能, 优于传统的鉴别特征提取方法。

关键词: 数据可视化; 图表示; 特征提取; 二次映射; 特征选择; 遗传算法

中图分类号: TP181; TP391

文献标识码: A

文章编号: 1004-731X (2009) 16-5080-04

Distinguishing Visual Feature Extraction Method Using Quadratic Map and Genetic Algorithm

WANG Jin-jia^{1,2}, LI Jing^{2,3}, ZHANG Tao^{1,2}, HONG Wen-xue²

(1. College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China;

2. College of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China;

3. College of Science, Yanshan University, Qinhuangdao 066004, China)

Abstract: The graphical representation of multi-dimensional data is a simplest method of the data visualization with high dimension. The visual barycentre graphical feature was proposed based on the radar plot of multi-dimensional data. The radar plot was involved with the feature order, which led that visual feature was involved with the feature order. The quadratic map was used to obtain visual features of the all feature orders. The distinguishing visual feature selection method was proposed based on the improved genetic algorithm (GA). For some UCI dataset such wine, breast cancer and diabetes, the obtained best classification error of distinguishing visual feature of radar plot is 0%、1.61% and 20.7%, which is very promising compared to the previously reported classification methods, and is superior to that of traditional feature extraction method.

Key words: data visualization; graphical representation; feature extraction; quadratic map; feature selection; genetic algorithm

引 言

分类是所有智能系统面对纷繁复杂的传感器数据时, 从中提取出有意义信息时所采用的第一个关键的处理步骤。在一个模式分类系统中, 一个理想的特征提取方法应该产生一种数据的表示^[1], 以使得后继的分类器变得简单。可以说特征提取在模式识别和机器学习领域占有十分重要的地位。一个常规的思路是提取数据的最有鉴别能力的特征, 这些特征对于类别信息无关的变换具有不变性, 例如线性判别分析(Fisher Linear Discriminant Analysis, FLDA)和广义线性判别分析(Generalized Discriminant Analysis, GDA)^[2]。

这里我们从信息可视化的角度重新审视特征提取问题。多元数据图表示是高维数据可视化的最简单的一种处理方法, 它包括平行坐标图、雷达图、散点图、脸谱图和星座图等^[3]。一个高维数据的样本的雷达图表示, 不仅实现了高维数据的可视化, 同时给了我们可视化的信息特征, 使我们更容易分类数据和理解数据。例如在合理安排好脸谱图的脸谱

特征和高维数据的特征变量的对应关系后, 脸谱图常用于基于人的视觉的直观分类, 它的本质是同一类的数据具有相似的脸谱, 不同类的数据具有不同的脸谱^[4-5]。从视觉上区分不同类的脸谱图, 其实质应该是多元数据的脸谱图图表示给出了某种具有可视化的和鉴别能力的特征。

国际模式识别协会的主席 Duin 指出^[1]：“在数据建模和数据预测上性能好的数据表示是一个挑战性的工作, 下一步工作应该是表示要既基于专家知识又能从数据中学习或优化。” Duin 和他的合作者提出了基于不相似度的表示方法, 并取得了成功^[6]。2006 年的 Nature 上发表论文指出^[7]：“果蝇能够利用两个图形在重心高度、朝向、大小和颜色等参数方面的不同。进行图形的分辨, 并对图案形成记忆。”这篇论文并回答了“基于这种图形识别及记忆形成的生物学结构基础是什么”。因此我们提出采用多元数据图表示的手段, 并基于^[7]提出的生物学基础提出了图形重心图形特征, 较完全地实现上述国际上的难题。

因此基于多元数据图表示, 我们提出了一种具有可视化和鉴别能力的特征提取新方法。以雷达图为例说明, 它首先将高维数据表示为雷达图的形式, 再从雷达图中提取可视化的图形特征, 然后送入分类器。由于雷达图表示与数据的特征排序有关, 导致图形特征深受特征排序影响, 所以基于二

收稿日期: 2008-04-25 修回日期: 2008-06-10

基金项目: 国家自然科学基金面上项目(60504035), 燕山大学优秀博士

生科学基金
作者简介: 王金甲(1978-), 男, 河南人, 讲师, 研究方向为模式识别及其应用; 洪文学(1953-), 男, 黑龙江人, 博导, 研究方向为图表示。

次映射,计算出所有特征排序下的图形特征,再从中基于遗传算法选择出具有鉴别能力的图形特征。葡萄酒、乳腺癌和糖尿病等 UCI 真实数据集的实验结果证实了我们的想法。

1 数据可视化

多元数据图表示是最简单的一种高维数据可视化技术,而雷达图是最常用的一种多元数据图表示方法。雷达图简单、主观,深入讨论它对于研究高维数据可视化技术很有帮助。下面以一个样本 $r = \{r_1, \dots, r_i, r_{i+1}, \dots, r_d\}$ 为例说明雷达图的作图方法,它是 $[0, 1]$ 均匀分布归一化样本。

雷达图的作图方法如下。(1) 先划一个单位圆,圆心代表 0,圆周代表 1;(2) 设原始数据共有 d 个变量,将圆周 2π 弧度等分为 d 个部分,并画出这 d 个半径,相邻半径之间的角度用弧度 $w_i = 2\pi/d$ ($i=1, 2, \dots, d$) 来表示。这 d 个半径就作为 d 个变量的坐标轴,第一维坐标轴一般表示 0 度的半径,第二维坐标轴表示 $360/d$ 度的半径,依次类推;(3) 将任一样本的归一化特征的 d 维变量值 r_i ($i=1, 2, \dots, d$) 按照次序画在 d 个均匀刻度的半径上,第 1 维变量值画在 0 度的半径上,第 2 维变量值画在 $360/d$ 度的半径上,依次类推。(4) 将所有半径上的 d 维变量值 r_i 所产生的点依次连接起来得到一个 d 多边形,这就产生了一个样本的雷达图。如图 1 所示。

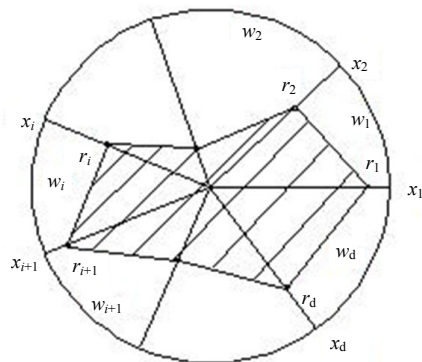


图 1 高维数据的雷达图表示的示意图

2 可视化特征提取

从图 1 上看,雷达图围成了 2 维平面里的一个封闭的有多个三角形组成的不规则多边形。每个三角形都是由相邻变量组成。从物理角度看,一个样本的雷达图中的每一个三角形都产生一个重心。从几何上讲三角形三条中线相交于一点,这个交点叫做三角形的重心,而且三角形的重心与顶点的距离等于它与对边中点的距离的 2 倍。如图 2 所示,已知三角形 AOB 的两边和一夹角, G 表示重心, OG 表示原点 O 到重心 G 的距离。我们定义雷达图的可视化特征为重心特征,它是原点到重心的距离,一个 d 维特征的样本产生一个 d 维重心特征的样本的公式如下。

$$G_i = \sqrt{\frac{r_i^2 \sin^2 \omega_i}{3} + \left(\frac{r_i \cos \omega_i - \frac{r_{i+1}}{2}}{3} + \frac{r_{i+1}}{2} \right)^2} \quad (1)$$

$i=1, \dots, d$

从(1)式可以看出, G_i 的值与 r_i, r_{i+1} 和角度 w_i 有关,而 $w_i = 2\pi/d$ ($i=1, 2, \dots, d$) 是与 d 有关的一个定值。当 r_i, r_{i+1} 互换位置时, G_i 的值不变,即满足交换律 $f(r_i, r_{i+1}) = f(r_{i+1}, r_i)$ 。特殊的,当 $i=d$ 时, $i+1=d+1$ 是第 $d+1$ 维,由于雷达图是沿着圆周排列变量,可以认为第 $d+1$ 维就是第 1 维。总之,重心可视化特征提取方法是一种无导师的非线性映射方法。

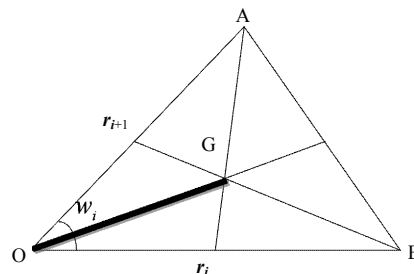


图 2 雷达图表示的可视化重心特征的示意图

多元数据图表示与数据的特征排序有关。传统图表示方法要么指定特征排序要么就是默认的原始特征排序。很明显,数据的特征排序一定会影响雷达图的形状,从而影响雷达图可视化特征的提取,最终必然导致分类性能受影响。即可视化特征提取方法不是具有鉴别能力的特征提取方法。

3 二次映射

二次映射是指将一个 d 维数据 $x = \{x_1, \dots, x_i, x_{i+1}, \dots, x_d\}$ 映射到一个新的 $d(d+3)/2$ 维的空间里,如公式(2)所示。显然 $x_i x_j = x_j x_i$, 公式(2)的下三角矩阵就忽略了,否则不能忽略。

$$Y = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \\ x_1 x_1 & x_1 x_2 & \cdots & x_1 x_d \\ & x_2 x_2 & \cdots & x_2 x_d \\ & & \ddots & \vdots \\ & & & x_d x_d \end{bmatrix} \quad (2)$$

如果我们把 $x_i x_j$ 替换为一般的两个变量表达的一个函数 $f(x_i, x_j)$, 那么我们比较容易与公式(1)建立起联系。可视化特征 $f(r_i, r_{i+1})$ 可以更一般地表示为 $f(r_i, r_j)$, i 和 j 分别表示是第 i 和 j 维。由于 $f(r_i, r_j)$ 满足交换律,因此可以将 $f(r_i, r_j)$ 作为 $f(x_i, x_j)$ 的一种特殊的实现。即可视化特征作为二次映射矩阵 Y 的元素,因此所有可能的特征排序下的雷达图的可视化重心特征都是 Y 的元素。这样上节提到的特征排序影响可视化特征的问题就能够解决,它转化为一个特征选择问题,即将一组 d 维数据按照可视化特征提取公式(1)通过二次映射升维到 $d(d+3)/2$ 维的空间,然后从 $d(d+3)/2$ 维的高维空间中选择出 d 维可视化特征,并希望这 d 维可视化特征具有类别鉴别能力。

4 基于遗传算法鉴别可视化特征提取方法

遗传算法(Genetic algorithms, GA)是对生物系统里达尔

文的自然选择和遗传理论进行直接模拟的一个自适应优化搜索算法, 优于传统的启发式搜索方法。GA 工作在称为一个种群的一组有效解上, 基于达尔文的适者生存的原则, 在一系列的迭代遗传操作后, GA 获得了优化解。最终的种群中的适应度最高的染色体对应的问题的可接受的解。GA 能有效处理较大的搜索空间, 因此比其它方法如模拟退火算法、Tabu 搜索算法更易得到优化解。

将遗传算法用于特征选择^[8], 关键的问题是如何将问题的解编码为染色体。一般选择简单的二进制编码, 我们选择了一种有效的编码方式: 整数编码。它要求染色体中的基因是从 1 到特征最大维数 $d(d+3)/2$ 间互不相同的整数, 基因间排列顺序任意且基因总长度等于选择的特征维数 d 。多个随机选择的特征染色体组成了初始种群, 并对种群进行标准的选择、交叉和变异等遗传操作。针对特征选择问题, 标准的遗传操作可能产生不可行解。一种解决方法是修改选择、交叉和变异等遗传操作以产生可行解。另一种解决方法是对标准的遗传操作后的染色体进行后处理。常见的后处理包括拒绝方法、修补方法和惩罚方法。拒绝方法抛弃遗传过程产生的所有不可行解, 它是最简单也是效率最低的方法。修补方法将得到的不可行解通过修补变为可行解, 创建修补过程相对较容易。惩罚方法是最常用的方法, 从本质上讲, 这种方法通过对不可行解的惩罚将约束问题转化为无约束问题, 任何对约束的违反都要在适应度函数中添加惩罚项。大部分惩罚方法都是依赖于问题的有参数的具有可变惩罚率的方法, 设计惩罚函数没有一般的指导性原则。而且惩罚方法通常并不拒绝每代中的不可行解, 但这一点就不适合我们的问题。基于以上分析, 我们选择了保持标准的遗传操作, 并提出了适合我们问题的基于修补方法的后处理, 我们称为修补算子, 它将标准的遗传操作产生的所有不可行解修补为可行解。然后估计每个修补后的合法的染色体的适应度函数。

如何将问题的解编码为染色体是遗传算法使用中的关键问题。另外也需要考虑当染色体需要解码成为解时从基因型空间到表现型空间的映射关系, 以及当染色体被遗传算子操作时的变形问题。我们的算法对于传统的 GA 有三个主要修正: 染色体的设计, 适应度函数, 和提出新的修补算子。最后给出基于遗传算法的特征选择方法的系统结构框图。

4.1 染色体设计

首先随机生成若干染色体的一个种群, 每个染色体都是一种特征选择。编码方案要求每个基因互不相同, 且必须是小于或等于最大维数的正整数, 并且基因长度等于选择的特征维数。为简单起见, 染色体的生成采用随机方法。随机方法是指随机产生一组数并取整数, 其中必须避免重复基因。

4.2 适应度函数

适应度函数表示染色体的质量, 常用的适应度函数有距离测度、信息熵测度、相关测度、一致性测度和分类器准确

率等。这里适应度函数设计为每个染色体对应的选择的鉴别可视化特征的分类准确率。分类性能采用分类器的 10 倍交叉验证的准确率。适应度函数值越高, 说明选择的鉴别可视化特征越好。适应度函数采用分类器的 10 倍交叉验证(10 CV)的准确率, 分类器选用线性判别分析(linear discriminate analysis, LDA)分类器, 即等协方差的正态分布概率型的分类器, 最近邻分类器和 K 近邻分类器(K-nearest neighbors, KNN), 和支持向量机(support vector machines, SVM)。

4.3 遗传操作和修补算子

适应度函数评估了染色体的质量。通过评估适应度函数选择染色体进行繁殖。选用轮盘赌的选择操作, 即染色体的适应度函数越高被选择进交配池的概率越高。交叉和变异函数是影响适应度函数的主要操作, 而交叉率和变异率的影响有限。一般交叉操作采用单点交叉或多点交叉交换两个染色体上的基因。它是一个很重要的遗传操作, 允许在搜索空间中挖掘新的解区。但这种传统方法经常导致交叉后的染色体不符合特征选择编码方案的要求。一般变异操作改变染色体中的某个基因和某些基因。这也经常导致变异后的染色体不符合特征选择编码方案的要求。为了应用传统的遗传操作而不做修改, 我们提出了一种新的遗传后处理操作: 修补算子。修补算子包括四步: 第一步, 对染色体进行取整操作; 第二步统计重复基因的位置(即第二次或第二次以后出现的基因), 并统计大于最大维数和小于 1 的非法基因的位置; 最后随机选择没有出现的基因放置到第二步统计的基因位置处并保证修补后的染色体一定是合法的。子代采用了精英保留策略, 最终组成新的种群以便下次进行迭代遗传操作。进化过程执行很多代直至到达收敛条件为止。

4.4 基于二次映射和遗传算法的鉴别可视化特征提取方法

图 3 给出了基于二次映射和遗传算法的鉴别可视化特征提取方法的示意图, 总结的主要步骤如下。

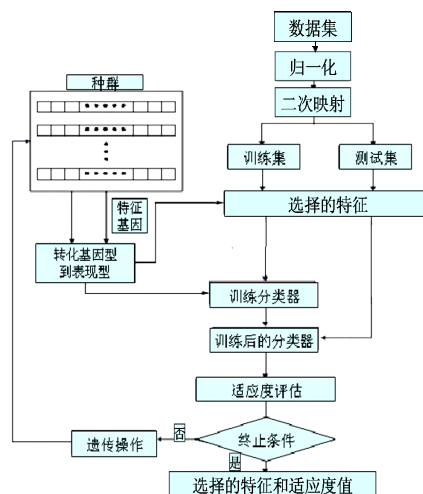


图 3 基于二次映射和遗传算法的鉴别可视化特征提取方法

- (1) 数据归一化。归一化的主要优点是避免较大数值范围的属性影响较小数值范围的属性,另一个优点是避免计算中的数值困难。一般地,每一个特征线性归一化到[0 1]。选用这种归一化的原因是雷达图作图的需要。
- (2) 二次映射:按照可视化特征提取公式(1)将 d 维数据进行二次映射得到 $d(d+3)/2$ 维的新数据。
- (3) 产生初始种群,每一个染色体表示选择一组 d 维特征,即从 $d(d+3)/2$ 维选择出 d 维。
- (4) 初始种群的适应度数评估:基于初始染色体所选择的特征,将新数据分成训练样本集和测试样本集。训练集用来训练分类器,测试集用于计算分类正确率。分类正确率即为某种染色体选择特征的适应度函数。
- (5) 遗传操作:这一步通过选择、交叉、变异、修补和子代替换父代等遗传操作搜索更好的染色体,即更好的解,更好的特征选择。
- (6) 遗传种群的适应度数评估:基于遗传得到的染色体所选择的特征,重新将新数据分成训练样本集和测试样本集。训练集用来训练分类器,测试集用于计算分类正确率。
- (7) 当终止条件如最大迭代次数满足时,过程就结束;否则返回 5)遗传操作。

5 仿真结果和讨论

为了评估提出的鉴别可视化特征的分类正确率,我们对从 UCI^[9]下载的 3 个数据集进行了实验。这些数据集经常用于文献中各种模式识别任务的性能比较。表 1 总结了这些数据集的样本数、特征数和类别数。

表 1 从 UCI 数据库选择的数据集			
数据名字	样本数	特征数	类别数
IC1 wine	178	13	3
Breast-cancer-Wisconsin	683	9	2
Pima Indians diabetes	768	8	2

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,分类器错误率的估计采用 10 CV,取 10 次独立实验的平均结果。10 CV 是指将数据集样本共分成 10 份,使用 (10-1)份样本设计分类器,并估计剩余的 1 份样本;这个过程重复 10 次就完成了一次 10 CV 的实验。

归一化方法采用常见的[0,1]线性归一化方法。作为比较,我们测试了原始特征排序下的可视化特征的分类性能;还测试了传统的 FLDA 和 GDA 特征提取方法的分类性能。所有特征提取方法产生的维数都为原始特征维数。对于选择核参数为 4 的径向基核函数的 GDA 方法,这可能处理复数问题,我们简单地取实部。遗传算法特征选择的参数为个体选择 20 个,最大迭代次数是 50 代,选择方法为轮盘赌,交叉方法为单点交叉,交叉率为 1,变异方法为均匀变异,变异率为 0.02。LDA 和 KNN 分类器采用 PRTOOLS 工具箱^[10]。KNN 我们使用了最近邻(1NN)和 3 近邻(3NN)。SVM 采用

LIBSVM 工具箱及其提供的 Matlab 接口^[11]。SVM 我们采用了核参数为 0.5 的径向基函数核,超参数 C 选择 100。基于遗传算法选择的鉴别可视化特征的维数等于原始特征维数。

我们的实验运行在 Matlab R2006a 环境上,计算机环境是 Intel Pentium III 2.4 GHz CPU 和 512MB RAM。

表 2 到表 4 给出了这些数据集的 10 次 10 CV 平均分类错误率实验结果,方差都在 0.01 以下。所有特征下,每个分类器的最优分类错误率用黑体标出。从表中可看到,鉴别可视化特征的分类性能基本上优于可视化特征和 FLDA 和 GDA 特征提取方法。GDA 特征的核参数没有优化,但分类性能也很有竞争力。可视化特征和 FLDA 特征的性能较差。

文献[12]报道了几乎所有常见的分类器在一些数据包括我们使用的三个数据的分类性能。我们也给出了最新的研究进展如文献[6]的不相似度的表示方法的分类结果。表 5 给出了这些数据集的文献[12]和文献[6]报道的最佳分类性能和我们的鉴别可视化特征最佳分类错误率性能比较。结果表明我们的方法全部胜出。

表 2 Wine 数据集的 10 次 10 CV 平均分类错误率				
	LDA	1NN	3NN	SVM
FLDA 特征	0.0219	0.1758	0.1955	0.0337
GDA 特征	0.0045	0.0062	0.0056	0.0056
可视化特征	0.0112	0.0618	0.0618	0.0393
鉴别可视化特征	0.0042	0.0054	0.0090	0.0000

表 3 Breast-cancer-Wisconsin 数据集的 10 次 10 CV 平均分类错误率				
	LDA	1NN	3NN	SVM
FLDA 特征	0.0307	0.0565	0.0577	0.1362
GDA 特征	0.0250	0.0366	0.0321	0.0293
可视化特征	0.0395	0.0439	0.0264	0.0293
鉴别可视化特征	0.0322	0.0205	0.0161	0.0249

表 4 Pima Indians diabetes 数据集的 10 次 10 CV 平均分类错误率				
	LDA	1NN	3NN	SVM
FLDA 特征	0.2303	0.3096	0.2721	0.2643
GDA 特征	0.2238	0.3069	0.2775	0.2266
可视化特征	0.2266	0.3333	0.2865	0.2301
鉴别可视化特征	0.2070	0.2617	0.2122	0.2174

表 5 3 个数据集的鉴别可视化特征的最优分类错误率性能和文献[6, 12]的 10 次 10 CV 平均分类错误率性能比较			
数据名字	本文方法	文献[6]	文献[12]
IC1 wine	0.0000	0.000	0.011
Breast-cancer-Wisconsin	0.0161	0.002	0.025
Pima Indians diabetes	0.2070	0.211	0.224

6 结论

基于多元数据图表示的数据可视化方法,提出了可视化特征提取方法。基于二次映射,提出了鉴别可视化特征提取的思想,基于改进的遗传算法的特征选择实现了鉴别可视化特征提取。3 种 UCI 数据集的实验结果表明鉴别可视化特征的最佳分类错误率分别达到了 0%、1.61%和 20.7%,优于报

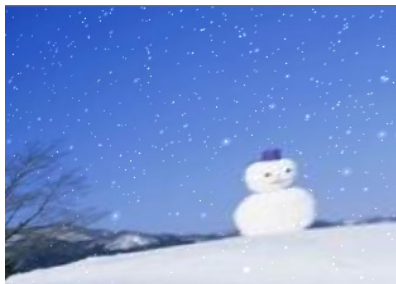


图 3(d) -3°C 时飘雪模拟情况

5 结论

不规则自然景观的模拟是一项非常复杂的课题,本文所提出的基于温度变化以及物理模型来模拟雪景的方法是从真实性角度出发,尽量考虑到自然界的真实情况,达到了真实性和实时性的平衡,因此该方法是合理的也是可行的。图 4 是真实飘雪场景的照片,对比可以看出本文模拟的效果还是基本真实的。由于只侧重于温度这个因素,其它更细微、更繁琐的因素仍未考虑完全。同时模拟风场变化对飘雪的影响仍然是未来待研究的重点。



图 4 真实飘雪场景照片

(上接第 5083 页)

道的常用的分类性能,优于传统的鉴别特征提取方法,优于可视化特征。本文的创新在于基于二次映射将鉴别可视化特征提取问题转化为了特征选择问题。

未来工作包括新的可视化特征提取方法,遗传算法的特征选择中的涉及的关键参数值的选择策略的分析和不同参数值对结果影响的试验,遗传算法的计算效率(或复杂度)分析和比较,基于其它特征选择方法的鉴别可视化特征提取。

参考文献:

- [1] Robert P W Duin, E Pekalska. The Science of Pattern Recognition: Achievements and Perspectives [C]// W. Duch, J. Mandziuk (eds.), Challenges for Computational Intelligence, Studies in Computational Intelligence. Germany: Springer, 2007, 63: 221-259.
- [2] Cheong Hee Parka, Haesun Park. A comparison of generalized linear discriminant analysis algorithms [J]. Pattern Recognition (S0031-3203), 2008, 41(3): 1083-1097.
- [3] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005.
- [4] Astel A. Classification of drinking water samples using the Chernoff's Faces Visualization Approach Polish [J]. International Journal of Environmental Studies (S0020-7233), 2006, 15(5): 691-697

参考文献:

- [1] Nishita T. A Modeling and Rendering Method for Snow by Using Metaball [J]. Computer Graphics Forum (S0167-7055), 1997, 16(3): 357-364.
- [2] Paul Fearing. Computer Modelling of Fallen Snow [Z]. Canada: University of British Columbia, 2000.
- [3] Premoze. Geospecific Rendering of Alpine Terrain [R]. Spain: Proc the 10th Eurographics Workshop on Rendering, 1999.
- [4] 陈彦云, 孙汉秋, 郭百宁. 自然雪景的构造和绘制[J]. 计算机学报, 2002, 25(9): 916-922.
- [5] Ohlsson. Real-time Rendering of Accumulated Snow [Z]. Sweden: Uppsala University, 2004.
- [6] M S Langer. A Spectral-particle Hybrid Method for Rendering Falling Snow [Z]. Canada: McGill University, 2004.
- [7] Ingar Saltvik. Parallel Methods for Real-time Visualization of Snow [D]. Norway: Norwegian University of Science and Technology, 2006.
- [8] 王长波, 王章野, 夏添, 等. 飘雪场景的实时模拟[J]. 软件学报, 2004, 15(Special Issue): 278-286.
- [9] 王润杰, 田景全, 倪政国. 基于粒子系统的实时雨雪模拟[J]. 系统仿真学报, 2003, 15(4): 495-501.
- [10] 徐利明, 姜显明. 基于粒子系统与 OpenGL 的实时雨雪模拟[J]. 计算机仿真, 2005, 22(7): 242-245.
- [11] Wes Junker. Winter Weather Forecasting [EB/OL]. (2000-6) [2008-4]. www.hpc.ncep.gov/html/hpcframes.html.
- [12] 王顺领, 李学庆, 王善斌. 虚拟场景中的实时降雪模拟[J]. 计算机技术与发展, 2006, 16(3): 152-157.
- [13] T B Moeslund. Modeling Falling and Accumulating Snow [Z]. Denmark: Computer Vision and Media Technology, Aalborg University, 2005.
- [5] 王金甲, 洪文学. 一种 K-均值脸谱图聚类新算法[J]. 仪器仪表学报, 2007, 28(10): 1916-1920.
- [6] E PeRkalska, Robert P W Duin, Pavel Paclik. Prototype Selection for Dissimilarity-based Classifiers [J]. Pattern Recognition (S0031-3203), 2006, 39(2): 189-208.
- [7] Liu Gang, Seiler Holger, Wen Ai. Distinct memory traces for two visual features in the Drosophila brain [J]. Nature (S0028-0836), 2006, 439(7076): 551-556.
- [8] Cheng-Lung Huang, Chieh-Jen Wang. A GA-based feature selection and parameters optimization for support vector machines [J]. Expert Systems with Applications (S0957-4174), 2006, 31(2): 231-240.
- [9] Asuncion A, Newman D J. UCI Machine Learning Repository Irvine, CA, USA, University of California, School of Information and Computer Science.. [DB/OL]. (2007) [2009-7-25]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [10] R P W Duin, P Juszczak, P Paclik, E Pekalska, D de Ridder, D MJ Tax. PRTools4, A Matlab Toolbox for Pattern Recognition [K]. The Netherlands: Delft University of Technology, 2004.
- [11] Chih-Chung Chang, Chih-Jen Lin. LIBSVM: a library for support vector machines, Software [Z/OL]. (2001) [2009-7-25]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [12] Datasets used for classification comparison of results. [DB/OL]. (2004) [2008-03-23]. <http://www.is.umk.pl/projects/datasets.html>.