# 结合二进制 PSO 的 SVM 在园林评价中的应用

李 雪,刘 弘
LI Xue,LIU Hong

山东师范大学 信息科学与工程学院,济南 250014
School of Information Science and Engineering,Shandong Normal University,Jinan 250014,China
E-mail:sdjnlx1024@163.com

**LI Xue,LIU Hong.Binary PSO combination of SVM in evaluation of garden.Computer Engineering and Applications, 2009,45(6):213-215.**

**Abstract**: According to the SVM algorithm theory,SVM based garden design evaluation model is established.By introducing binary PSO algorithm,the characteristic parameters of garden design is chosen to address the dimension disaster caused by a large number of irrelevant or redundant features,the garden design evaluation is maken by SVM multi-classifiers.Case analysis shows that this method can enhance the accuracy and reliability of the garden design evaluation.

**Key words**: Particle Swarm Optimization(PSO);Support Vector Machine(SVM);feature selection;garden design evaluation;multi-classifiers

**摘 要**:根据支持向量机算法的原理,建立基于支持向量机的园林设计评价模型,通过引入二进制微粒群算法对影响园林设计的特征参数进行选择,解决了大量无关或冗余特征所造成的"维数灾难"和降低分类器性能的问题,利用 SVM 多类分类器实现了对园林设计的评价。实例分析表明,该方法提高了园林设计评价的准确性和可靠性。

**关键词**:微粒群算法;支持向量机;特征选择;园林设计评价;多类分类器
**DOI**:10.3778/j.issn.1002-8331.2009.06.061 **文章编号**:1002-8331(2009)06-0213-03 **文献标识码**:A **中图分类号**:TP391

## 1 引言

园林规划设计需要满足布局合理性、适用性、美观性等各方面的要求,评价一个园林设计的好坏,需要考虑多个因素,相应的需要多个评价指标。本文使用 SVM 对园林设计进行评价,虽然它可以有效地处理高维数据,但是实验表明在训练 SVM之前,应该考虑进行特征子集选择[1]。Barzilay 和 Brailovsky 等人在特征选择之后使用支持向量机作为分类器,识别效果得到了显著的提高[2]。Cao 和 Tay 等人使用支持向量机对金融数据进行预测,在训练支持向量机之前使用 GA 选择特征子集,获得了很好的效果[3]。任江涛和赵少东等人提出在训练支持向量机之前,使用二进制 PSO 算法进行特征选择并同时进行参数优化,也取得了较好的分类效果[4]。因此把二进制微粒群算法引入支持向量机进行特征选择,选出那些少量的非常有用的特征参数作为评价因子,这样做有效地降低了分类的错误率,提高了评价的准确性和可靠性。

## 2 园林设计评价指标体系

园林设计需要满足布局合理性、美观性、适用性等多方面的要求,每个方面又存在着复杂的联系,为此,许多学者对大量园林设计进行了深入研究:智强、杨梅等人在《景观设计概论》,安秀在《公共设施与环境艺术设计》书中,提出了布局合理性方面的各个规范标准;卢新海、杨祖达在《园林规划设计》书中提出了适用性方面的规范标准;以及李铮生在《城市园林绿地规划与设计》书中提出了美观性方面的标准。根据这些规范标准建立了园林设计的评价指标体系,将指标体系分为三大主要方面,总共包含 25 个评价因子,其内容如下:

(1)布局合理性($M_1$)。评价因子包括:园林中山水布局、道路布局、停车场位置($p_1$)、楼体位置($p_2$)、标志性建筑的位置($p_3$)等 10 个子项[5-6]。

(2)适用性($M_2$)。评价因子包括:建筑容积率($p_4$)、道路覆盖率($p_5$)、娱乐设备的数量($p_6$)等 6 个子项[7]。

(3)美观性($M_3$)。评价因子包括:绿地占有率($p_7$)、水面占地率($p_8$)、植物种类($p_9$)、亭子种类($p_{10}$)、湖中是否有岛等 9 个子项[8]。

实验将通过对各评价因子的评估得出三个方面的得分,再根据这三个方面的评价得分得出整个园林设计的评价结果。

**基金项目**:国家自然科学基金(the National Natural Science Foundation of China under Grant No.69975010,No.60374054,No.60743010);山东省自然科学基金(the Natural Science Foundation of Shandong Province of China under Grant No.Y2003G14,No.Z2006G09)。

**作者简介**:李雪(1985-),女,硕士研究生,主要研究方向为进化计算、计算机辅助设计;刘弘(1955-),女,博士,教授,主要研究方向为 CSCW、多 Agent 系统、机器学习。

**收稿日期**:2008-01-07 **修回日期**:2008-03-26

## 3 支持向量机和微粒群算法的简介

### 3.1 支持向量机理论介绍

假设存在训练样本$\{x_i, y_i\}, i=1,2,\cdots,l, x_i \in R_m, y_i \in \{-1,+1\}$，$l$ 为样本数，$m$ 为输入样本维数。在线性可分的情况下就会有一个超平面，使两类样本完全分开，超平面为：

$$\omega x_i + b = 0 \tag{1}$$

式中：$\omega$ 为权值向量；$b$ 为分类阈值。

求解最优超平面就是对给定的训练样本，找到权值 $\omega$ 和阈值 $b$ 的最优值，使得权值代价函数最小化，即：

$$\min\varphi(x)=\frac{1}{2}\|\omega\|^2$$

$$\text{s.t.}\quad y_i(\omega x_i+b)\geqslant 1, i=1,2,\cdots,l \tag{2}$$

当训练样本线性不可分时，需引入松弛变量 $\xi_i$ 和惩罚参数 $C$。则超平面优化问题为：

$$\min\varphi(x)=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}\xi_i$$

$$\text{s.t.}\quad y_i(\omega x_i+b)-1+\xi_i\geqslant 0, i=1,2,\cdots,l \tag{3}$$

$$\xi_i \geqslant 0$$

将每一个样本点用一个非线性函数 $\varphi(x)$ 映射到高维特征空间，再在高维特征空间进行线性回归。

对式(1)引入拉格朗日函数得到：

$$\varphi'=\frac{1}{2}\|\omega\|^2+\sum_{i=1}^{l}\alpha_i[y_i(\omega x_i+b)-1+\xi_i] \tag{4}$$

解此方程不易求解，可解此方程的对偶方程如下：

$$\max Q(a)=\sum_{i=1}^{l}a_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_i a_j y_i y_j K(x_i,x_j) \tag{5}$$

相应的分类决策函数为：

$$f(x)=\text{sgn}[\sum_{i=1}^{l}y_i a_i^* K(x_i,x_j)+b^*] \tag{6}$$

应用 SVM 模式进行分类的基本思路可概括为：首先将输入向量映射到一个特征空间，然后在特征空间中寻找优化的线性分界线，即构建一个可分离两类的超平面，使两类正确分开。SVM 的训练过程就是寻找全局最优解。

### 3.2 PSO 算法简介

Kennedy 和 Eberhart[9]受到鸟群捕食行为的研究结果启发，于 1995 年提出微粒群优化(PSO)算法。PSO 算法具有执行速度快、受问题维数变化影响小等优点，迅速得到了人们的重视。算法描述如下：

设搜索空间为 $D$ 维，总微粒数为 $n$。第 $i$ 个粒子位置表示为向量 $\boldsymbol{X}_i=(x_{i1},x_{i2},\cdots,x_{iD})$；第 $i$ 个微粒迄今为止搜索到的最优位置为 $P_i=(p_{i1},p_{i2},\cdots,p_{iD})$，也称 $pbest$，整个微粒群迄今为止搜索到的最优位置为 $P_g=(p_{g1},p_{g2},\cdots,p_{gD})$，也称 $gbest$，第 $i$ 个微粒的位置变化率(速度)为向量 $\boldsymbol{V}_i=(v_{i1},v_{i2},\cdots,v_{iD})$。每个微粒的位置按如下公式进行变化("飞行")：

$$v_{id}(t+1)=v_{id}(t)+c_1\times rand\times(p_{id}(t)-x_{id}(t))+ c_2\times rand\times(p_{gd}(t)-x_{id}(t))$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1) \tag{7}$$

$$1\leqslant i\leqslant n \quad 1\leqslant d\leqslant D \tag{8}$$

其中，$c_1$、$c_2$ 为正常数，称为加速因子；$rand$ 为[0,1]之间的随机数，称惯性因子。

为了适应 PSO 算法在离散问题中的应用，Kennedy 和 Eberhart[10]于 1997 年提出了二进制 PSO 算法。在二进制 PSO 算法中，每个粒子被编码为一个二进制向量。在二进制粒子中，速度定义了粒子的每个位置赋值为 1 的概率，因此要通过转换函数将速度转换到区间[0.0,1.0]。本研究中采用 sigmoid 函数。二进制的粒子更新公式如下：

$$p_{ij}=\begin{cases}1, rand()s(v_{ij})\\ 0, \text{otherwise}\end{cases} \tag{9}$$

为了获得更好的优化效率和效果，本研究对 PSO 算法进行了一些改进，在每轮迭代中将适应度最差的 10%粒子进行变异。实验证明这 10%的粒子浪费计算资源，而进行变异后粒子群能够更快地找到最优值，并且避免了容易聚集于局部最优的情况。

### 3.3 使用二进制 PSO 进行特征选择的几个重要步骤

(1)编码：根据特征选择的特点，把每一个特征定义为粒子的一维二进制变量，变量长度就等于所有特征的数量，如果第 $i$ 位为 1，那么第 $i$ 个特征就被选中，否则这个特征就被屏蔽。因此，每一个粒子就代表了一个不同的特征子集，也就是一个候选解。

(2)初始化：初始化种群就是随机产生一组粒子，然而通过这种方式获得的种群中每一个个体的"1"或者"0"的数量是大致相同的，也就是每个个体的特征数量大致相同。为了获得不同数量的特征，采用的方法是首先随机产生每个粒子所含"1"的个数，然后再把这些"1"随机分布在这个粒子的所有维中。实验证明这种方法更能有效地反映出特征的多样性。

(3)适应度的评价：特征子集选择的目标是使用少量的特征达到相同或更好的分类效果，因此适应度的评价包含两部分内容：①验证的准确率。使用特征子集中确定的特征来训练分类器，用交叉验证结果来评价分类器的性能。②使用的特征数量。每个特征子集包含一定数量的特征，如果两个特征子集获得的准确度相同，包含特征比较少的子集就被选中。在准确率和特征数量这两个因素中，需要重点考虑的是准确率，因此适应度函数确定为如下形式：

$$fitness=10^4\times(1-Accuracy)+Ones \tag{10}$$

式中 $Accuracy$ 是每个个体获得的准确率，其中使用的是 SVM 分类器 5 阶交叉验证的结果作为准确率的值，$Ones$ 是粒子中包含"1"的数量。这里为了提高准确率的重要性，把准确率的权值定为 10 000。高准确率就意味着适应度值小，该特征子集就有可能在竞争中获胜。

## 4 园林设计评价模型

利用支持向量机进行园林设计评价的基本原理为：根据园林设计评价体系，将每个评价参数作为输入，园林设计评价等级作为输出，通过支持向量机寻求它们之间的非线性关系。

支持向量机算法实际上是一个两类分类器，而园林设计评价有多个级别，两类分类器不能满足评价的要求。为此，需要利用基于 SVM 的多类分类器，需要利用解决多类分类问题的方法[11]，目前存在两种方法，一种是"一对多"方法，另一种是"一对一"方法，本文采用的是"一对一"的方法。

多类分类问题可以如下描述：根据给定训练集

$$T=\{(x_1,y_1),\cdots,(x_l,y_l)\}\in(X\times Y)^l,$$

其中 $x_i\in X=R_n, y_i\in Y=\{1,2,\cdots,K\}, i=1,2,\cdots,l$。

寻找一个决策函数 $f(x): X=R_n \to r$ 由此可见，求解多类分

类问题，实质上就是找到一个把 $R_n$ 上的点分成 $k$ 部分的规则。“一对一”方法(One-against-One Method)：这种方法也是基于两类问题的分类方法，不过这里的两类问题是从原来的多类问题中抽取的。具体做法是：分别选取两个不同类别构成一个SVM子分类器，这样共有 $k(k-1)/2$ 个 SVM 子分类器，在构造类别 $i$ 和类别 $j$ 的 SVM 子分类器时，在样本数据集中选取属于类别 $i$、类别 $j$ 的样本数据作为训练数据，并将属于类别 $i$ 的数据标记为正，将属于类别 $j$ 的数据标记为负。“一对一”方法需要解决如下的最优化问题：

$$\min_{\omega^{ij},b^{ij},\varepsilon^{ij}} \frac{1}{2}(\omega^{ij})^{\mathrm{T}}\omega^{ij}+C\sum_t \varepsilon_t^{ij}$$

$$\text{s.t.}(\omega^{ij})^{\mathrm{T}}\varphi(x_t)+b^{ij}\geqslant 1-\varepsilon_t^{ij},\ \text{if}\ y_t=i$$

$$(\omega^{ij})^{\mathrm{T}}\varphi(x_t)+b^{ij}\geqslant -1+\varepsilon_t^{ij},\ \text{if}\ y_t=j$$

$$\varepsilon_t^{ij}\geqslant 0 \qquad (11)$$

解决这一优化问题后，即用训练样本进行训练后就可以得到 $k(k-1)/2$ 个 SVM 子分类器。测试时，将测试数据对 $k(k-1)/2$ 个 SVM 子分类器分别进行测试，并累计各类别的得分，选择得分最高者所对应的类别为测试数据的类别。在这种方法中，需要多个两类问题的分类器。对 $k$ 类问题，就有 $k(k-1)/2$ 个两类分类器，这种方法的优点是其训练速度较快。本文中园林设计评价系统共分为 3 级：Ⅰ为优秀，Ⅱ为良好，Ⅲ为较差，只需要 3 个分类器，属于简单分类，所以采用训练速度较快的”一对一”方法。

## 5　实验分析

实验的运行环境为 Windows XP，开发环境为 Visual C++ .NET，并结合使用了 Matlab7.0 科学计算软件。学习算法：SVM 采用径向基核函数，惩罚参数 $C$=1 000，利用“一对一”分类方法；二进制 PSO 参数设置为：$c_1=c_2$=2.0，粒子数量为 30，算法停止条件为迭代 100 次。

经二进制 PSO 特征选择之后，原有的 25 个评价因子最终剩下 10 个，即 $p_1$~$p_{10}$ 这 10 个因子，比以前没有进行特征选择时减少了一半多，这样大大提高了 SVM 的分类效率。

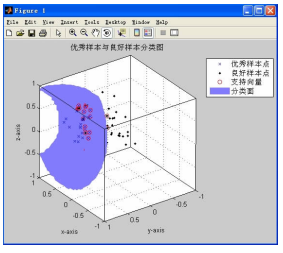
通过对现有园林设计方案的研究，从中选择了 100 个设计方案，并根据所有的评价因子对每个园林设计方案进行专家人工打分，每个评价因子得分范围是[0，100]，分为三个等级：较差[0，25]，良好[26，75]，优秀[76，100]。利用综合模糊评价法得到园林三个大方面($M_1$、$M_2$、$M_3$)的得分，进而得到园林整体设计的评价结果：-1(较差)，0(良好)或者 1(优秀)。为了使实验结果更好地呈现，把专家人工打分的结果数据转化到[-1，1]。

对 25 个评价因子的得分利用综合模糊评价法可以得到园林三个大方面及整体设计的评价结果，这样就可以得到 100 组数据，其中 50 组数据作为训练样本即样本 $A_1$(表略)，50 组数据作为测试样本即样本 $A_2$。接下来，对进行特征选择后的 10个评价因子的得分利用综合模糊评价法也可以得到相应的三个方面及整体设计的评价结果，相应的得到 100 组数据，50 组作为训练样本即 $B_1$，数据见表 1，另 50 组数据作为测试数据即 $B_2$。然后调用 SVM 多分类器，对样本 $A_1$ 与 $A_2$，$B_1$ 与 $B_2$ 分别进行训练和测试。

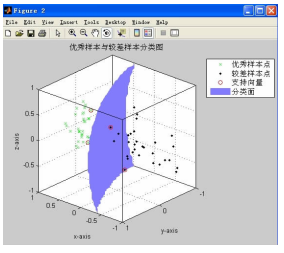
第一步，训练阶段。进行特征选择后的训练样本 $B_1$ 的训练结果如图 1 所示，进行训练后可以得到三个 SVM 子分类器，其中第一个分类器用于优秀样本与良好样本的分类(图 1(a))，第二个分类器用于优秀样本与较差样本的分类(图 1(b))，第三个分类器用于良好样本与较差样本的分类(图 1(c))。图中蓝色的部分代表超平面，·和×分别代表两类不同的样本，红色○代表支持向量机。经过 SVM 训练后，所有样本被超平面划分为三个等级。

**表 1　训练样本($B_1$)**

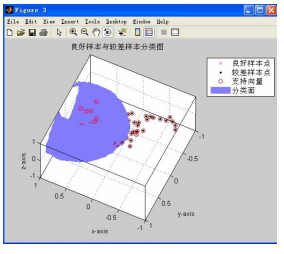
| 园林序号 | $M_1$ | $M_2$ | $M_3$ | 评价结果 |
|---|---|---|---|---|
| 1 | 0.95 | 0.94 | 0.93 | 1 |
| 2 | 0.85 | 0.79 | 0.92 | 1 |
| 3 | 0.40 | -0.10 | -0.32 | 0 |
| 4 | -0.31 | 0.12 | 0.34 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 48 | -0.67 | 0.82 | -0.56 | -1 |
| 49 | -0.05 | -0.76 | -0.89 | -1 |
| 50 | 0.92 | 0.47 | 0.76 | 1 |




(a)优秀样本与良好样本分类图




(b)优秀样本与较差样本分类图




(c)良好样本与较差样本分类图

**图 1　SVM 训练结果图**

训练样本 $A_1$ 经过训练后也会得到三个训练结果图(图略)，其训练结果和 $B_1$ 的相近，只是超平面的形状和位置发生了变化，本文不做详细介绍。

第二步，测试阶段。用上面两组测试样本分别对各自训练好的 SVM 分类器进行测试，实验结果如表 2 所示，没有使用特征选择的 SVM 多分类器含有 25 个全部的特征，分类错误率为 3.23%，使用二进制 PSO 特征选择之后，特征数量减少了 15 个，SVM 多分类器最优的分类错误率降低到了 2.35%，降低了 27.2%，说明经过特征选择之后，分类准确率有了提高，并且特征数量也大大减少。

**表 2　实验测试结果**

| 特征选择方法 | 特征数量 | 最优特征子集 | 错误率 |
|---|---|---|---|
| 无 | 25 | 全集 | 3.23% |
| PSO-SVM | 10 | 1110010100000110001110000 | 2.35% |

## 6　结束语

利用支持向量机方法构建园林设计分类评价模型，实现了对园林设计三种等级的智能评判，在训练前引入二进制 PSO进行特征选择，提高了 SVM 分类准确率。利用二进制 PSO 和支持向量机方法相结合进行评价，尽量剔除了评价的主观性，比专家的主观评判更符合客观实际，因此增强了结果的准确性、可靠性和客观公正性。

(下转 239 页)