

# 基于信息融合和 CS-SVM 的学生综合能力 评估方法的研究

杨亭榆, 傅成华

(四川理工学院自动化与信息工程学院, 四川 自贡 643000)

**摘要:**当代社会学生综合能力的好坏直接影响学生未来的发展,然而目前对学生综合能力评估仍旧存在一定的困难,针对这种情况提出了一种基于信息融合和 CS-SVM(布谷鸟优化的支持向量机)的学生综合能力评估方法。该方法首先将学生综合素质相关的数据融合成 SVM 的输入样本数据,然后再将此数据通过训练好的 CS-SVM 进行综合评估。MATLAB 仿真结果表明该预测结果与实际结果误差很小,能够在一定程度上评判出学生的综合能力。

**关键词:**学生能力;综合评估;信息融合;CS-SVM

**中图分类号:**TB115

**文献标志码:**A

## 引言

当今社会日新月异,随着科学的飞快发展,我国急需一批综合素质较强、具备独挡一面能力的应用型人才。目前很多用人单位去各个高校招聘时只能通过查看学生的考试成绩以及获奖证书来判定学生的优异度,这样尤其突出了高校对学生综合素质能力评判标准的重要性。高等教育多元化,各学科专业培养目标不同,因而需具备的能力素质不同,这就要求高校适时地加强对学生的综合素质培养力度,既要培养学生具备较强的专业知识,又要培养学生适应社会的竞争力,还要培养学生的团队凝聚力。因此需要一种方法对学生的综合素质进行评估,以期准确地对学生进行评优,从而正确

地反映学生的综合素质能力。

学生的考试成绩只能反映出学生的学习态度以及目前的学习状况,并不能体现一个人的综合素质能力,学生综合能力包含了考试成绩、思想道德素质、身心健康、动手操作、科学创新以及创新创业能力等多项能力<sup>[1-3]</sup>。传统的评价方法大多采用线性方法,人为的设定各项指标的权重进行线性计算,具有较强的主观性;模糊综合评价法<sup>[4-5]</sup>对各项指标进行模糊处理,具备较强的人为经验,虽然是一种非线性处理方法,但是不具备学习能力;神经网络评价法<sup>[6-7]</sup>虽然具备学习能力,但是不能对一些模糊信息进行有效处理。

以上评估方法互有优缺点,大多高校通常采用其中的一种方法进行评判。对此,本文采用人工智能的方

收稿日期:2017-10-31

基金项目:四川省科技厅支撑项目(2013GZ0030)

作者简介:杨亭榆(1993-),女,四川乐山人,硕士生,主要从事智能控制方面的研究,(E-mail)1274720592@qq.com

傅成华(1958-),男,四川富顺人,教授,主要从事先进控制与系统优化、神经网络与非线性处理等方面的研究,(E-mail)1920266248@qq.com

法,通过模拟实际人工操作处理方式,提出一种基于信息融合的CS-SVM学生综合能力评估方法。该方法与其他方法相比在一定程度上减小了人为主观因素,增强了评价信息的完整性,对学生各方面的能力进行了更综合更全面的融合,科学快速并较为精确地对学生综合能力进行分类评估。此方法设计步骤为先将各科学业成绩、活动创新成绩、实践实验成绩、毕业实习成绩、综合测评成绩等数据融合为SVM的输入样本,再将样本放入根据经验知识训练好的CS-SVM进行分类评估。其通过多信息融合,结合CS-SVM优秀的分类能力进行分类,这样能更全面的反应出学生的实际综合素质能力。

## 1 信息融合

信息融合是将多维信息进行分析和处理的方法和理论,通过将事物不同特征维度数据进行融合和处理,使得其结果能更加全面可靠。目前存在大量的融合算法,大致可分为三类:随机式算法、最小二乘法算法以及人工智能式算法。随机式算法包含多贝叶斯估计法、证据推理法、多级递归算法等;最小二乘法算法包含加权平均法、卡尔曼滤波法等优化算法;人工智能算法包含粗糙集推理、模糊推理、神经网络、支持向量机等<sup>[8]</sup>。伴随着时代的进步与发展,人工智能化时代已悄然来临,其中信息融合已逐渐变得更加智能并且能将多种具有不同特征的信息集成化。信息融合按照数据的抽象层次进行划分可分为数据级层次的融合、特征级层次的融合以及决策级层次的融合。数据级层次的融合,具有数据损失量少,精确度高但其对数据资源要求极其严苛;决策级层次的融合数据丢失量最大、精度最低;特征级层次的融合有数据损失且融合性能较低,介于数据层融合与决策层融合二者之间<sup>[9-10]</sup>。

本文利用特征级融合采集思想道德水平(A1)、专业修养水平(A2)、身心修养水平(A3)、人文修养水平(A4)以及能力修养水平(A5)等能表明学生能力的特征数据,融合成输入矩阵,最终通过SVM卓越的分类能力作出相应的决策。其中思想道德水平包括政治修养、道德修养、法律意识水平;专业修养水平包括专业理论成绩、实

验实践成绩、毕业实习成绩;身心修养水平包括心理素质水平和身体素质水平;人文修养水平包括自然科学知识、人文社科知识、文化艺术知识的修养水平;能力修养水平包括组织管理能力、学术研究能力、科技创新能力、艺术创新能力等。分别将以上多种能力利用加权平均法融合成能表征该类能力的数(A1-A5),使得评价信息完整化。其中加权平均法<sup>[11]</sup>可以用以下表达式表示:

$$A_i = \sum a_i \cdot S_i \quad (1)$$

其中, $A_i$ 为第*i*个指标加权后的分数, $a_i$ 为第*i*个指标信息融合中的权重, $S_i$ 为第*i*个指标的分值。

$$a_i = \frac{1 - H_i}{m - \sum_{i=1}^m H_i} \quad (2)$$

其中, $0 \leq a_i \leq 1, \sum_{i=1}^m a_i = 1$ 。

$$H_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij}, i = 1, 2, 3, \dots, m \quad (3)$$

$$f_{ij} = r_{ij} / \sum_{j=1}^n r_{ij}, k = 1/\ln n \quad (4)$$

其中, $m$ 为评价指标, $n$ 为评价对象。

## 2 CS-SVM 算法

### 2.1 SVM 算法

SVM是人工智能领域能解决非线性数据分类的无监督机器学习算法。与传统的非线性分类模型相比,人工神经网络(ANN)<sup>[12]</sup>是一种多感知器(Perceptron)相结合的方法来解决非线性分类问题,它在很大程度上依赖于学习速率、隐含层结构和节点数目等一系列参数,参数的好坏会极大影响神经网络的分类效果。而支持向量机的关键是利用核函数将低维空间中难以分类的向量集映射到高维空间,建立分类超平面。将核函数的非线性问题的数据集转化为核空间中的线性可分离数据集,其中基于最大裕度思想的支持向量机只需要少量的参数就能进行调整<sup>[13-14]</sup>。

就数据二分类问题而言,假设存在一个训练样本集 $(x^{(i)}, y^{(i)})$ ,其中 $i = 1, \dots, n$ , $x^{(i)}$ 是样本特征, $y^{(i)}$ 是样本类型, $n$ 代表样本编号,分类决策方程可表示为:

$$f(x^{(i)}) = w^T x^{(i)} + b \quad (5)$$

由此可得到超平面函数方程:

$$w^T x^{(i)} + b = 0 \quad (6)$$

其中,  $w$  代表超平面法向量,  $b$  代表偏置量。自变量  $x^{(i)}$  则可以通过核函数映射到高维空间, 利用  $w^T x^{(i)} + b > 0$  或  $w^T x^{(i)} + b < 0$  来判别  $x^{(i)}$  属于哪一类。最优分类面如图 1 所示。

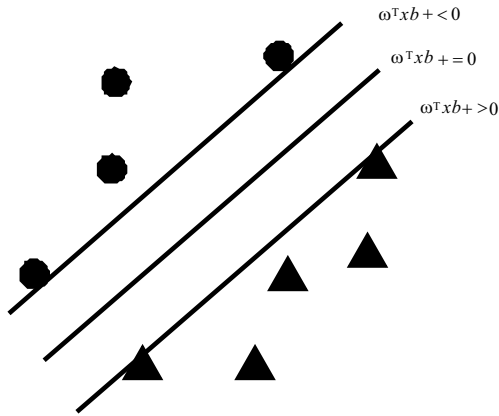


图 1 最优分类面

在实际数据分类应用中用函数间隔来表示点  $X$  到超平面的距离,“间隔”越大,分类可信度越高。得出较大化分类间隔器,可以将其表示为:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \quad (7)$$

$$\begin{aligned} \text{s. t. } & y_i (w^T x_i + b) \geq 1 - \varepsilon_i \\ & \varepsilon_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (8)$$

其中,  $C$  是惩罚因子,是控制目标函数取得最大超平面以及最小偏差量权重的参数,  $\varepsilon_i$  为松弛变量,表示函数误差。

将线性不可分数据原始问题转化为对偶问题,其可以用函数表示为:

$$w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} \alpha_i \alpha_j (x^{(i)}, x^{(j)}) \quad (9)$$

式中,  $\alpha_i$  为拉格朗日乘子,通过计算将其转化成最优分类函数:

$$f(x) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b \right) \quad (10)$$

其中,  $K(x_i, x_j)$  为 SVM 的核函数,核函数是连接低维空间与高维空间的桥梁,常见的核函数包括多项式核、径向基核、Sigmoid 核、Mercer 核,本文采用径向基核函数<sup>[15]</sup>。

## 2.2 CS 算法

布谷鸟是一种巢寄生繁殖的鸟类,它将自己的蛋产在别的鸟类的窝中,让其他鸟类代为孵化和育雏。CS 算法是由剑桥大学 YangXS 和 DebS 于 2009 年提出的一种模拟布谷鸟寄生寻窝繁殖、具有启发性思维的智能算法<sup>[16]</sup>。该算法是通过将布谷鸟随机寻窝下蛋的过程与鸟类的 Levy Flight 方式相结合形成。CS 算法繁殖下一代的环境背景可定义为<sup>[17-18]</sup>:

(1) 种群内布谷鸟每次产卵量为 1,即可假设该卵为最佳解,该最优解在固定的可选鸟窝范围内进行随机选择。

(2) 选择出的最佳鸟窝和最佳解被保存至下一代。

(3) 布谷鸟产于其他鸟窝内的卵被鸟窝主人发现存在一定的概率,因而布谷鸟需要寻找新的鸟窝。

(4) 布谷鸟的 Levy Flight 方式可表示为如下函数:

$$S_{g+1,i} = S_{g,i} + \vartheta \times \frac{\text{rand}_u}{\text{rand}_v^{1/\beta}} \times \chi(\beta) \times (S_{g,i} - S_{best}) \quad (11)$$

其中,  $S_{g,i}$  为第  $i$  个鸟窝第  $g$  代参数值,  $S_{g+1,i}$  为执行 Levy Flight 方式后的参数值,  $S_{best}$  为最佳参数,  $\vartheta$  为步长控制系数,  $\text{rand}_u$  和  $\text{rand}_v$  为服从标准正态分部的随机数,  $\chi(\beta)$  的大小取决于  $\beta$  值大小,  $\chi(\beta)$  通常取 0.6667。由式(10)可知,当  $S_{g,i}$  取  $S_{best}$  时,  $S_{g,i}$  与  $S_{g+1,i}$  相等,即当前取得最优参数。

SVM 对数据进行分类主要取决于  $C$  与  $g$  的大小。在本文中  $C$  值的大小将影响学生能力样本的学习;  $g$  为核函数参数,其值的大小将影响样本投射空间的分布复杂度。本文利用 CS 算法对  $C$  与  $g$  参数进行寻优处理。详细寻优步骤如下:

(1) 对 CS 算法相关参数以及各参数取值范围进行初始化。本文假设布谷鸟数量为 30,外来卵被发现的概率取 0.3,繁殖迭代次数取 300 代。

(2) 布谷鸟第一次产卵后,通过计算取出适应度最佳的鸟窝再将其保存给下一代。即计算出 30 组  $C$ 、 $g$  参数进行交叉计算从而验证出正确率,分别取出具有最大正确率的那组  $C$ 、 $g$  参数保留起来以便与下一次进行对比。

(3) 布谷鸟通过 Levy Flight 方式飞行,寻找新的鸟窝,将其和旧鸟窝比较,选出最佳的那组窝。即利用 Levy Flight 处理原来的  $C$ 、 $g$  参数以获取新的  $C$ 、 $g$  值,通过将产生的新的  $C$ 、 $g$  值与之前的  $C$ 、 $g$  值进行对比,选择出最佳  $C$ 、 $g$  值。

(4) 鸟窝主人发现布谷鸟蛋的概率为 0.3,此时一部分布谷鸟就需要找寻新的鸟窝。即需要对 30 组  $C$ 、 $g$  参数随机分配一个 0~1 的数,对小于 0.3 的  $C$ 、 $g$  参数随机重置。

(5) 布谷鸟再次将新窝与旧窝进行对比,优胜劣汰。即将各个  $C$ 、 $g$  参数的交叉正确率进行比较并再次获取新的  $C$ 、 $g$  参数。再回到第(3)步,直到满足迭代要求获得最佳正确率以及最优  $C$ 、 $g$  参数。

### 3 实验结果与分析

在实际生活中,各高校对学生的成绩等级的划分参照的依据不同,本文将成绩等级划分为 5 个区间段,分别是  $[90, 100]$ 、 $[80, 90]$ 、 $[70, 80]$ 、 $[60, 70]$ 、 $[0, 60]$  然后将其划分为优秀、良好、中等、及格、不及格 5 个等级。按此评价指标将能表现学生综合素质能力的的数据经过信息融合处理来训练 CS-SVM。根据以上分类依据,选取 20 组不同等级的训练样本以及测试样本,作为训练和测试 CS-SVM,让其具备智能化的分类评定经验。将学生能力水平划分成 5 种类型,并分别定义 0 为优秀、1 为良好、2 为一般、3 为及格、4 为差。将某大学毕业生四年来能体现学生综合素质能力的各项成绩融合成数据样本输入至 CS-SVM 使其得出一个分类结果,其中通过信息融合处理后的部分样本数据见表 1。

表 1 部分样本数据

A1	A2	A3	A4	A5	类别
92.0259	89.3269	91.7369	87.5268	92.3560	0
88.1525	80.3335	85.0310	79.7180	82.6525	1
73.7369	71.2106	70.3118	72.6201	60.4253	2
67.3118	60.0297	65.4213	63.2217	56.7324	3
56.3317	35.3127	57.4891	60.4427	37.7469	4

将该样本数据通过 SVM 进行交叉验证,通过交叉验证可以看出参数  $C$ 、 $g$  的值不同得到的正确率也存在

差异,本文利用布谷鸟算法多次验证找出了最大化正确率下对应的参数  $C$ 、 $g$ , 然后将其植入 SVM 中。图 2 是利用布谷鸟算法交叉验证取得的最优正确率。

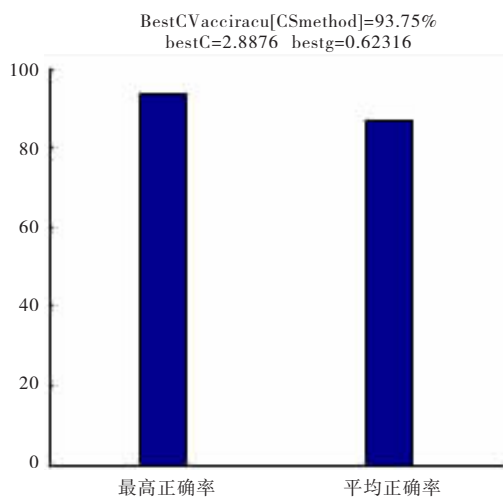


图 2 CS 交叉验证正确率

由图 2 可知,通过布谷鸟算法进行交叉训练获得的最佳正确率为 93.75%,其中  $C$  取 2.8876,  $g$  取 0.62316,然后将 20 组待测样本数据通过训练好的 CS-SVM 中进行分类测试,以达到检验 CS-SVM 准确度的目的。

图 3 为利用 CS-SVM 对待测样本进行测试的验证图,根据图 3 可得出其分类正确率为 87.5%,由此可知 CS-SVM 算法基本可以满足对学生综合能力的分类要求。

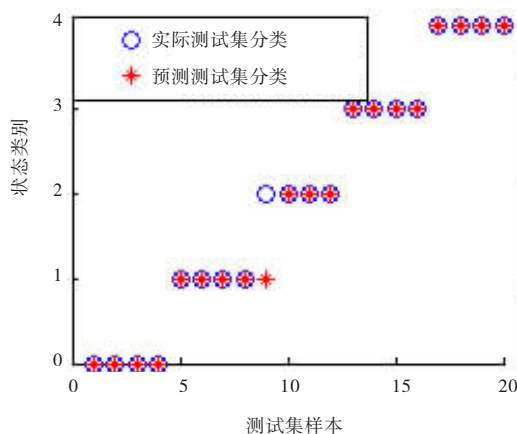


图 3 CS-SVM 测试样本预测

在不使用优化算法直接用多分类的支持向量机对学生综合素质能力进行分类,随机选取核函数以及惩罚因子,同样将能体现学生能力的上述数据样本直接通过

SVM进行训练,然后对测试样本进行预测验证,得到的分类结果如图4所示。由图4可知,未通过优化算法直接训练测试得到的分类结果很差,其分类正确率仅有43.2%,很难正确地将学生综合能力进行分类判定。

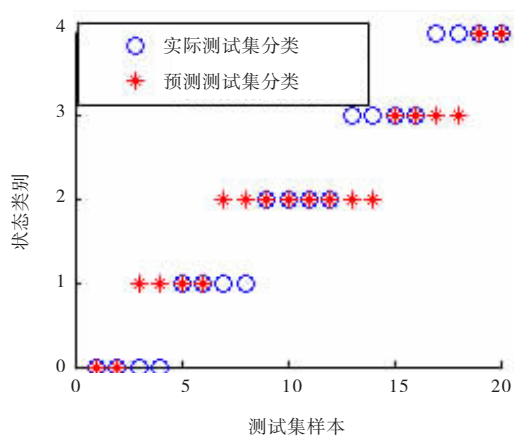


图4 无优化 SVM 测试样本预测

通过将不使用优化算法的SVM与使用CS-SVM算法对学生综合素质能力进行分类获得的分类正确率进行比较,显然CS-SVM算法更为可靠。再将此分类结果与该校毕业生工作几年的状况进行对比分析可知,通过CS-SVM算法获得的分类结果中成绩优秀者往往都成为了部门经理或者自己成为了老板,成绩良好者则大多小有成就,成绩中等者大多处于稳定状态,成绩及格者大多仅能解决温饱问题且为月光族,成绩差者仍旧处于四处找工作但四处碰壁整日劳累奔波的状态。以上验证了该方法能够较为精确地对学生综合能力进行评估,且实用性较好。

#### 4 结束语

本文提出的基于信息融合和CS-SVM的学生综合能力评估方法,将能够体现学生综合素质能力的各项成绩提取其特征数据,然后再将其进行信息融合,以此作为训练样本以及测试样本,进而利用CS算法对惩罚因子C和核函数g优化,使得SVM拥有较强的分类经验,最终利用此CS-SVM来对学生综合素质能力进行预测分类。通过实例仿真以及毕业生就业调查结果可以得出此算法能够在一定程度上对学生能力进行分类判别,从而较为准确地对学生的综合能力进行评定。

#### 参考文献:

- [1] 王桂芳.基于人工智能的大学生综合素质评价研究[D].北京:北京服装学院,2010.
- [2] 商利华.模糊聚类分析算法在学生成绩综合评估中的应用[J].电脑知识与技术,2014,10(5):1024-1026.
- [3] 刘幸,高延春.大学生综合能力评价体系与评价方法研究[J].西部素质教育,2015,1(8):33,49.
- [4] 智德.航海技术专业学生适任能力综合评估研究[J].电脑知识与技术,2015,11(22):176-177.
- [5] 刘敏慧.模糊综合评价法在核心素养评价中的应用研究[J].教育参考,2016(6):36-39.
- [6] 宣国庆.基于神经网络交叉覆盖算法的学生成绩预测[D].合肥:安徽大学,2011.
- [7] 张新亮.BP神经网络在高校学生综合素质评价中的应用[C]//第5届教育教学改革与管理工程学术年会论文集.重庆:重庆大学出版社,2012.
- [8] 谢振南.多传感器信息融合技术研究[D].广州:广东工业大学,2013.
- [9] 化柏林,李广建.大数据环境下多源信息融合的理论与应用探讨[J].图书情报工作,2015,59(16):5-10.
- [10] 李会民,马桂英.异构多源信息融合方法概述[J].嘉应学院学报,2016,34(2):21-26.
- [11] 高超.浅析加权平均法在多目标决策中的应用[J].电脑知识技术,2010,6(16):4495-4496.
- [12] 段玉三.人工神经网络文献综述[J].科技风,2011(5):185.
- [13] 房汉鸣,税爱社,汪辉,等.支持向量机动态多分类方法[J].后勤工程学院学报,2017,33(2):90-96.
- [14] 刘东启,陈志坚,徐银,等.面向不平衡数据分类的复合SVM算法研究[J/OL].计算机应用研究,2018,35(4).(优先出版).
- [15] 安旭,张树东.基于支持向量机的模糊特征分类算法研究[J].计算机工程,2017,43(1):237-240,246.
- [16] YANG X S, DEB S. Engineering optimization by cuckoo search[J].International Journal of Mathemati-

- cal Modelling & Numerical Optimisation,2010,1(4): 330-343.
- [17] 黄继达.布谷鸟算法的改进及其应用研究[D].武汉:华中科技大学,2014.
- [18] 孙晨,李阳,李晓戈,等.基于布谷鸟算法优化 BP 神经网络模型的股价预测[J].计算机应用与软件,2016,33(2):276-279.

## A Study on the Evaluation Method of Students' Comprehensive Ability Based on Information Fusion and CS-SVM

*YANG Tingyu, FU Chenghua*

(School of Automation & Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China)

**Abstract:** The stand or fall of contemporary social students' comprehensive ability directly affect the future development of the students. However, contemporary society to the students' comprehensive ability evaluation still exist certain difficulties. For this kind of situation, a comprehensive assessment method for students based on information fusion and cs-svm (support vector machine for cuckoo optimization) was proposed. The students' comprehensive quality-related data are integrated into the input sample data of SVM firstly, and then the data through the trained CS-SVM are evaluated. The MATLAB simulation results show that the predicted results and actual results error is very small to a certain extent, so it can judge the students' comprehensive ability.

**Key words:** comprehensive ability; comprehensive assessment; information fusion; CS-SVM