高校教师教学质量评价体系与评价方法的创新研究

李 强　黄 婷　石 磊　叶曼莉

（南京农业大学 理学院，江苏 南京 210095）

摘　要：教学质量评价是反映高校老师课堂教学的重要指标，但是目前多数高校选用的教学质量评价体系和方法缺乏科学性，使得教学质量评价结果与实际情况有偏差，导致高校教师对教学投入的主动性和积极性不高。本文通过对目前教学质量评价体系的研究，提出一种创新的评价方法。
关键词：高校教学；教学质量；评价方法；支持向量机法
中图分类号：G642　　文献标志码：A　　文章编号：2096-000X（2015）12-0005-04

Abstract: Teaching quality evaluation is an important indicator of college classroom teaching quality. However, the teaching quality evaluation system and method used in the majority of colleges and universities are lacking in scientificity. It makes the results of teaching quality evaluation deviated. What is worse, the bias may lead to a negative attitude towards college classroom teaching. In this paper, an novel evaluation method is put forward based on the research of the current teaching quality evaluation system.
Keywords: college teaching; teaching quality; assessment method; support vector machine

一、研究背景

改革开放以来，我国高等教育取得优异成绩。但是在我国高等教育大众化的推进过程中，其中存在的质量问题也渐渐凸显出来。为了提高教学质量，很多学校选择进行教学质量评价。但是，目前评价指标体系缺乏合理性。对教师的教学质量评价通常是由学生、教学督导和各学院教学指导委员会评价三部分组成。

在教学质量评价体系不够合理的情况下，任课老师的积极性会受到打击，会出现教师“为评价而教学”、学生“为成绩而评教”的情况。因此，我们希望通过研究寻求一个更加合理的评价方法。

二、支持向量机方法

支持向量机(SVM)是由 Vapnik 领导的 AT&T Bell 实验室研究小组在 1963 年提出。1995 年 Cortes 和 Vapnik 首先提出比较完善的 SVM 方法[1]。支持向量机是建立在统计学习理论的 VC 维理论和结构风险最小化原理基础上的，根据有限样本信息在模型复杂性和学习能力之间寻求最佳折衷，以期获得最好的泛化能力。

支持向量机中的一大亮点是在传统的最优化问题中提出了对偶理论，理论和实际相结合的降低了原有的复杂度。

SVM 的关键在于核函数，核函数可以认为是构建和开启支持向量机的钥匙。矢量集的低维空间往往很难划分，解决的办法是考虑低维和高维空间的千丝万缕的联系。但这种方法的困难是，使计算复杂性超出正常范围，而为了解决了这个问题各种探索不断，其中核函数的出现让大家喜出望外。换言之，一旦选择恰当的核函数，高维空间分类功能就能实现。在支持向量机理论中，采用不同的核函数将导致不同的支持向量机算法。

选取满足模型的适当核函数后，由于确定核函数的已知数据也存在一定的误差，考虑到推广性问题，研究者们引入了松弛系数和惩罚系数[2]两个参变量来加以校正。在确定了核函数基础上，经过大量实验改变不同的参数取定这两个因素，在核函数中参数确定后再进行模型运行。当然误差是相对存在的，领域和学科相异导致对误差的要求不同[3]。

（一）支持向量机线性和非线性分类

支持向量分类(support machine classification-SVC)[4][5]是从线性可分情况下的最优分类面发展而来的，也是统计学习理论中最实用的部分，其基本思想可用图 1 的两维例子解释说明。图中，实心圆和空心正方形代表两类样本，分类超平面很容易找到且不唯一，分别过各类中离分类超平面最近的样本且平行于分类超平面的平面，它们之间的距离叫做分类间隔。

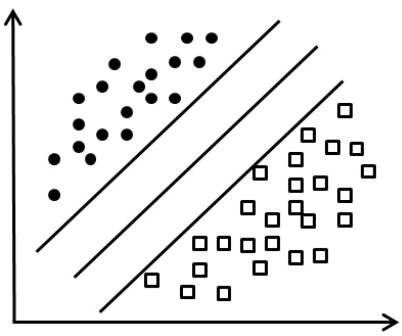

图 1 线性可分下的最优分类

作者简介：李强，副教授。

所谓最优分类超平面就是要求分类面不但能将两类正确分开,而且要使分类间隔最大。而距离最优分类超平面最近的向量称为支持向量(Support Vector,SV)[6]。

线性分类超平面的求解比较容易,但在很多数学问题中无法找到线性分类超平面,需要将其推广到非线性分类超平面中。例如对于图2(a)所示的问题,分别属于正类和负类,显然这时用线性分类超平面划分会产生很大的误差,这类问题称为线性不可分问题。这时就必须使用非线性分类学习机。

对于这类问题,显然不能用超平面去划分,那么人们自然希望用“超曲面”来代替超平面。对于超平面,支持向量机利用“间隔”最大的思想,去确定最优超平面,但超曲面没有间隔的概念。这里,通过一个映射 $\psi:x\rightarrow H$,把寻找超曲面的问题转化为寻找超平面的问题,将输入向量 x_i 映射到高维特征空间 H。

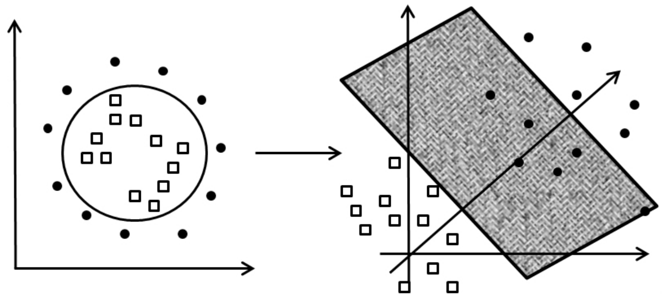

图2 线性不可分数据的非线性分类

（二）支持向量机分类数学原理

对于给定的训练样本集:$(x_1,y_1),(x_2,y_2),\cdots\cdots,(x_n,y_n)$

其中 $x_i\in R^N$,是N维向量,$y_i\in\{-1.1\}$或者 $y_i\in\{1,2,\cdots\cdots,k\}$,$i=1,2,\cdots,n$。在训练学习的过程中获得模型决策函数 $M(x)$,训练样本 $y_i=M(x)$,对于预测样本集 $x'_1,x'_2,\cdots,x'_m$,同样可以得到相对合理的 y_i 值。当 $y_i\in\{-1.1\}$时,称为二分类问题;当 $y_i\in\{1,2,\cdots,k\}$时,为k分类问题。

对于训练样本集是线性可分的情况,通过给定的训练样本集,求模型函数:$y=f(x)=sgn(w\cdot x)+b$

假设对整个样本数据 x_i 满足不等式:

$$\begin{cases} w\cdot x_i+b\geq 1 & yi=1 \\ w\cdot x_i+b\leq 1 & yi=-1 \end{cases}$$

通过合并后可得:$y_i(w\cdot x_i+b)\geq 1\quad i=1,2,\cdots,l$

已知点 x_i 到超平面的距离为:$d(w,b,x_i)=\frac{\|w\cdot x_i+b\|}{\|w\|}$

已知最优分类超平面的概念的基础上,分类间隔函数如下所示:

$$p(w,b)=\min d(w,b,x_i)+\min d(w,b,x_j)$$
$$=\min\frac{\|w\cdot x_i+b\|}{\|w\|}+\min\frac{\|w\cdot x_j+b\|}{\|w\|}=\frac{1}{\|w\|}+\frac{1}{\|w\|}=\frac{2}{\|w\|}$$

要使分类间隔达到最大,即求 $\max\frac{2}{\|w\|}$。从而构造最优分类超平面的问题就变为最小化问题。

令 $g(w)=\frac{1}{2}<w\cdot w>$

对于线性可分的训练集,对应的优化问题为

$$\min\frac{1}{2}<w\cdot w>\quad s.t.\quad y_i(w\cdot x_i+b)\geq 1 \tag{3.4}$$

引入拉格朗日算子 λ_i 有:

$$L(w,b,\lambda)=\frac{1}{2}<w\cdot w>-\sum_{i=1}^{l}\lambda_i[y_i(w\cdot x_i+b)-1]$$

拉格朗日函数依次对 w,b 求偏导,并方程等于零得:

$$\frac{\partial l(w,b,\lambda)}{\partial w}=w-\sum_{i=1}^{l}y_i\lambda_i x_i=0$$

$$\frac{\partial l(w,b,\lambda)}{\partial b}=\sum_{i=1}^{l}y_i\lambda_i=0$$

求得 $w=\sum_{i=1}^{l}y_i\lambda_i x_i$,$\sum_{i=1}^{l}y_i\lambda_i=0$

将它们带入拉格朗日函数中,上述分类问题的最优解为:

$$L(w,b,\lambda)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(x_i\cdot x_j)\quad s.t.\quad \sum_{i=1}^{l}y_i\lambda_i=0$$

得到最优解为 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_l^*)^T$

由KKT定理[7]可知最优解满足

$\lambda_i>0$

$\lambda_i[y_i(w\cdot x_i+b)-1]=o$

其对偶形式为:

$$\max\Phi(\lambda)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(x_i\cdot x_j)\qquad \lambda\geq 0$$

于是训练集的模型函数 $y=f(x)=sgn(w\cdot x)+b$ 变为:

$$f(x)=sgn(\sum_{i=1}^{l}y_i\lambda_i^*(x_i\cdot x_j)+b^*)$$

线性可分的样本下求得的决策函数存在且不唯一,在所有的决策函数中存在唯一最优的决策函数,但它的形式也不唯一,可能是直线,也可能为超平面,在理论上就寻找 w,b 的最优解。决策函数使得各类别的分类间隔最大,它的求解最终可转化为求解一个二次凸规划问题,依据最优化理论,该问题存在全局最优解[8]。

针对线性不可分的样本数据,利用松弛变量 ξ,一定程度放大约束条件,同时引入惩罚系数 C,在达到最少的错分样本和最大的分类间隔的前提下,将线性不可分的情况顺利转换为线性可分的情况[9]。

引入松弛变量 ξ_i,$y_i(w\cdot x_i+b)\geq 1$ 变为 $y_i(w\cdot x_i+b)\geq 1-\xi_i$

相对应的优化问题变为

$$\min\frac{1}{2}<w\cdot w>+C\sum_{i=1}^{l}\xi_i$$

$$s.t.\quad y_i(w\cdot x_i+b)\geq 1-\xi_i,\xi_i\geq 0$$

其中 w 为权向量,b 为偏置,ξ_i 为松弛变量。其中对样本的约束由惩罚因子 C 决定。

引入拉格朗日算子可得方程为

$$L(w,\xi,b,\lambda,\gamma)=\frac{1}{2}<w\cdot w>+C\sum_{i=1}^{l}\xi_i-\sum_{i=1}^{l}\lambda_i[y_i(w\cdot x_i+b)-1+\xi_i]-\sum_{i=1}^{l}\gamma_i\xi_i$$

上述拉格朗日方程分别对 w,b,ξ 求偏导，并结果等于零得：

$$w=\sum_{i=1}^{l}y_i\lambda_ix_i,C=\lambda_i+\gamma_i,\sum_{i=1}^{l}y_i\lambda_i=0$$

将它们带入拉格朗日函数中,上述分类问题的最优解为：

$$L(w,\xi,b,\lambda,\gamma)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(x_i\cdot x_j)$$

约束条件为：$\sum_{i=1}^{l}y_i\lambda_i=0$，$C\geq\lambda_i\geq0,i=1,2,\cdots,l$

得到最优解为：$\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_l^*)^T$

由 KKT 定理可知最优解满足

$\lambda_i,\gamma_i,\xi_i>0$

$\lambda_i[y_i(w\cdot x_i+b)-1]=0$

$\gamma_i\xi_i=0$

选择 λ^* 得到分量 $0<\lambda_j^*<C$

$$b^*=y_i=\sum_{i=1}^{l}y_i\lambda_i^*(x_i\cdot x_j)$$

其对偶形式为：

$$\max\Phi(\lambda)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(x_i\cdot x_j)\qquad\lambda\geq0$$

于是训练集的模型函数 $y=f(x)=sgn(w\cdot x)+b$ 变为：

$$f(x)=sgn(\sum_{i=1}^{l}y_i\lambda_i^*(x_i\cdot x_j)+b^*)$$

在解决非线性问题时，常用的数学方法是构造一个非线性映射函数，将原本样本空间映射到一个高维或者无穷维的特征空间，通过高维特征空间的线性支持向量机的方法构建最优超平面来解决低维样本空间中的非线性分类问题，不必知道非线性映射函数的显式表达式,通过选择适当的核函数,在高维的特征空间只需进行内积运算便可对应原样本空间的非线性输出[10]。常用的核函数有以下几种：

(1) Linear:$K(x,y)=x\cdot y$

(2) Potynomiat:$K(x,y)=[(x\cdot y)+1]^d$

(3) RBF:$K(x,y)=e^{-\|x-y\|^2/2\sigma}$

(4) Sigmoid:$K(x,y)=\tanh[v(x,y)+c]$

支持向量机针对的是二分类问题，对于多分类问题需要转化为二分类问题,一般常用的有两种转化方法："一对多"转换和"一对一"转换,"一对多"是通过把多分类问题转换为多个二分类问题来解决;"一对一"转换指构建多个分类机。

由线性支持向量机可知,可知二次规划的目标函数为：

$$L(w,\xi,b,\lambda,)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(\psi(x_i)\cdot\psi(x_j))$$

$$=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_jK(x_ix_j)$$

其中 $K(x_ix_j)=\psi(x_i)\cdot\psi(x_j)$,则 $\Phi(\lambda)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{l}y_iy_j\lambda_i\lambda_j(x_i\cdot x_j)$

于是训练集的模型函数为 $f(x)=sgn(\sum_{i=1}^{l}y_i\lambda_i^*K(x_i\cdot x_j)+b^*)$

三、运用支持向量机法评价高校教师教学质量的基本思路

(一) 数据筛选

规则 1：假设学院共有 W 个教师,方便起见不妨设每个教师为 N 个学生上课,M 全部评估指标的数目为 S_{kij},i 为学生对教师 k 的第 j 项评估指标的打分,$E_k(j)$和 $D_k(j)$分别为教师 k 的第 j 项评估指标的均值和方差。

$$E_k(j)=\frac{1}{N}\sum_{i=1}^{N}S_{kij};D_k(j)=\frac{1}{N}\sum_{i=1}^{N}(S_{kij}-E_k(j))^2$$

如果 $D_k(j)=0$,表明所有学生对该教师的第 j 项评估指标的打分都一样,则该评估数据将不被列为挖掘对象。

规则 2：E_k 和 D_k 为教师的总体评估统计指标：

$$E_k=\frac{1}{M}\sum_{j=1}^{M}E_k(j);D_k\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}(S_{kij}-E_k(j))^2$$

如果 $D_k=0$,表明所有学生对教师 K 的 M 个评估指标的数据完全一致,则该评估数据将被删除。

规则 3：$E_i(k)$和 $D_i(k)$分别是学生 i 对教师 k 的评教打分的均值和方差：

$$E_i(k)=\frac{1}{M}\sum_{j=1}^{M}S_{kij};D_i(k)=\frac{1}{M}\sum_{j=1}^{M}(S_{kij}-E_k)^2$$

如果 $D_i(k)=0$,表明学生 i 对教师 k 的所有评估指标的打分无差别,则该数据将不被选择。

规则 4：E_i 和 D_i 为教师评估指标的均值和方差：

$$E_i=\frac{1}{W}\sum_{k=1}^{W}E_i(k);D_i=\frac{1}{WN}\sum_{k=1}^{W}\sum_{j=1}^{M}(S_{kij}-E_i)^2$$

如果 $D_i=0$，表明学生 i 对所有教师的评估指标的评分没有差别,那么该数据将不会对数据挖掘产生积极作用,则不予选择。

(二) 数据处理

根据随机生成的数据,设定 A,B,C,D,E,F 是六个评价指标,分别为教学态度,教学目标,教学内容,教学方法,教学技能,教学效果。假设每个系都有 20 位老师,并根据化学数学物理的顺序依次编号。化学系教师总平均分 86,数学系教师总平均分 85，物理系教师总平均分 75。把 90 分及以上列为Ⅰ级,80-90 分为Ⅱ级,70-80 分为Ⅲ级,60-70 分为Ⅳ级。然后将不同学科内的分数根据公式 $\widetilde{X}_i=\frac{X_i-\min(X_i)}{\max(X_i)-\min(X_i)}$ 进行标准化。利用支持向量机方法进行教学质量等级分类。本文在 RStudio[11]平台下进行编程和运行,R 运行代码如下：

```
> library(e1071)
```

```
> edu <- read.table("edu.txt",header=T,row.names=1)
> obj <- tune(svm, types~., data = edu, ranges = list(gamma = 2^(-1:1), cost = 2^(2:4)),tunecontrol = tune.control(sampling = "fix"))
> obj
> library(kernlab)
> edumodel_RBF <- ksvm(types ~ ., data = edu, type = "C-bsvc", kernel = "rbfdot", kpar = list(sigma = 1), C = 4,prob.model = TRUE)
> predict(edumodel_RBF, edu[, -7], type = "probabilities")
> edunorm <- read.table ("edu_norm.txt",header=T,row.names=1)
> obj <- tune (svm, types~., data = edunorm, ranges = list(gamma = 2^ (-1:1), cost = 2^ (2:4)),tunecontrol = tune.control(sampling = "fix"))
> obj
> library(kernlab)
> edunormmodel_RBF <- ksvm (types ~ ., data = edunorm, type = "C-bsvc", kernel = "rbfdot", kpar = list(sigma = 0.5), C = 4,prob.model = TRUE)
> predict(edunormmodel_RBF, edunorm[, -7], type = "probabilities")
```

表 1 标准化后部分预测结果

教师编号	预测结果	实际结果
12	II	I
45	II	III
48	II	III
49	II	III
50	II	III
54	II	III
55	II	III
57	II	III
59	II	III

从计算结果数据可以看出，没经过标准化的数据，经过支持向量机分类，预测结果和实测结果一致。而标准化后的数据经过支持向量机分类，有部分结果发生变化，如表 1 所示，其中原来属于 I 类的一位化学老师标准化后被划入 II 类，8 位物理系的老师原来属于III类的，现在被划入 II 类。结果说明，利用支持向量机方法进行高校教学质量预测是可行的，并且数据处理是非常关键的。

四、结束语

对数据进行标准化处理并使用支持向量机理论后，不同类别不同教学方向的老师之间的评价结果具有了可比性，使得教学质量评价更具有公平性。完善后的体系不仅有利于教学管理部门加强对教师全面科学管理和准确了解教学情况，更能及时地将各种评价信息反馈给教师本人，促进教师的自我完善和发展。公平性的提高也使教师的教学积极性提高，从而进一步提高教学质量[12]。

参考文献

[1]V Vapnik and O Chapelle. Bounds on error expectation for support vector machines [M]. Neural Comput,2000（12）,9：2013-2036.

[2]R Stoean,M Preuss, C Stoean, E El-Darzi, and D Dumitrescu. An evolutionary approximation for the coefficients of decision functions within a support vector machine learning strategy [M].Foundations of Computational,Intelligence. 2009（1）：83-114.

[3]X Zhou,Y Xiong,WP Xiong, and LZ Xiong. Applied-information technology in fusion SVM partial least squares analysis of experimental data in the body of PET [C]. Advanced Materials Research,2014（951）：185-188.

[4]CW Hsu,CC Chang,and CJ Lin. A practical guide to support vector classification [M].Department of Computer Science. National Taiwan University,2003.

[5]B Waske, S Linden, JA Benediktsson, A Rabe, and P Hostert. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data [J]. Geosci. Remote Sens. 2010（48）,7：2880-2889.

[6]B Schölkopf and AJ Smola. Learning with kernels: Support vector machines,regularization,optimization,and beyond [M]. MIT press，2002.

[7]G Wanigasekara and T Fujimoto. Exposition of Karush-Kuhn-Tucker Theorem[C]. Annual Research Symposium,2014.

[8]LS Lasdon. Optimization theory for large systems [M]. Courier Corporation,2013.

[9]ML Huang，YH Hung，WM Lee，RK Li, and BR Jiang. SVM-RFE based feature selection and taguchi parameters optimization for multiclass SVM classifier[J]. Scientific World，2014.

[10]K Li, Z Huang,YC Cheng,and CH Lee. A maximal figure-of-merit learning approach to maximizing mean average precision with deep neural network based classifiers [C]. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014:4503-4507.

[11]JS Racine. RStudio: A platform-independent IDE for R and sweave[J].Applied Econometrics,2012（27）,1：167-172.

[12]王琰春，张胜军，赵菊梅. 高校教师教学质量评价存在的问题及对策[J].理工高教研究,2006（25）,1：66-67.