第60卷第6期 大连理工大学学报 Vol.60, No.6
2020年11月 Journal of Dalian University of Technology Nov. 2020

文章编号：1000-8608(2020)06-0599-11

基于支持向量回归的定性-定量因子混合建模方法

冯翔宇， 石茂林*， 马 跃， 宋学官， 孙 伟， 田 腾

（大连理工大学 机械工程学院，辽宁 大连 116024）

摘要： 在科学研究及工程实践中，输入参数经常同时包含定性因子与定量因子，为实现此类数据的有效建模，提出基于支持向量回归(SVR)的定性-定量因子建模方法，以用于工程实验及数值仿真的定性定量因子分析.引入超球面分解量化定性因子相关关系，构建了一种新型核函数描述定性因子与定量因子关联关系，提出了定性-定量因子支持向量回归算法实现定性定量数据的混合建模与预测.通过数值算例和经典工程算例，发现所提算法能提供相比于普通的支持向量回归算法及基于高斯过程回归的定性-定量因子算法更优的预测结果.以种植体骨应力分析为例，其中种植体材料类型为定性因子、结构参数为定量因子，实验结果表明所提算法能够显著提升骨应力预测精度，可为种植体的设计优化提供模型基础，揭示了所提算法的工程可用性.

关键词： 混合数据；核函数；支持向量回归；计算机实验

中图分类号： TP181 **文献标识码：** A **doi：** 10.7511/dllgxb202006007

0 引 言

数据建模是实现数据分析的基础，是完成工程系统设计、分析及优化的重要前提.在众多的数据建模方法中，支持向量回归(support vector regression，SVR)是具有代表性的一种，具有鲁棒性好、泛化学习能力强等优点[1-2]，因此在计算机数值仿真及工程设计中得到广泛应用.例如，Lute等[3]通过支持向量回归建立了桥梁设计参数与最大应力应变的映射关系，实现了斜拉桥设计主梁和塔柱的应力应变预测；Edwin Raja Dhas 等[4]基于支持向量回归建立焊接电学参数与焊接后部件残余应力相关关系，用于焊接过程的参数分析；Rahman 等[5]基于汽车悬架臂的荷载和材料数据，采用支持向量回归预测悬架臂最大位移量及最大主应力；Lostado 等[6]利用支持向量回归刻画了轴承预紧力、荷载和位移与车轴表面接触压力之间的关联关系.钱骥等[7]基于支持向量回归建立钢绞线应力预测的数学模型，并用于大跨度桥梁的检测监测；滕军等[8]以结构的动力和静力响应数据作为输入，多个设计参数作为输出，使用支持向量回归构建二者间的映射关系，实现了有限元模型多个参数的修正；狄勤丰等[9]基于有限元模型获得了套管部件的应力数据，通过支持向量回归建立了最大等效应力的预测模型.由以上工作可以看出，支持向量回归在预测、优化及分析等工程问题中均取得了成功应用.然而，以上研究中输入参数均为定量参数，而定量参数与定性参数并存的现象在工程实际中普遍存在[10].例如，机械零件尺寸设计参数可做定量改变，相应材料只能定性选择，设计出的零件应力应变等特性受尺寸定量参数与材料定性参数的共同影响[11-12].由于既有建模方法多基于输入参数间的空间距离描述彼此之间的相关关系，只能实现定量参数的数据建模，以至于既有方法不适用于描述定性参数间的相关关系.因此如何实现定量参数与定性参数的有效建模成为亟待解决的工程问题.

针对定量参数与定性参数混合建模问题，已经有很多学者提出了解决方案，并将定量参数及定性参数统称为定量因子(quantitative factors)和定性因子(qualitative factors)[13].其中具有代

收稿日期： 2020-03-21； **修回日期：** 2020-09-21.
基金项目： 国家重点研发计划资助项目(2018YFB1702500).
作者简介： 冯翔宇(1996-)，男，硕士生，E-mail：fxy@mail.dlut.edu.cn；石茂林*(1990-)，男，博士生，E-mail：shl5985336@mail.dlut.edu.cn.

表性的是基于高斯过程的定性-定量因子建模方法. Qian 等[14]建立了高斯过程模型中定性因子相关结构的一般框架,该框架采用半定规划估计法,保证了相关结构的正定性. Zhou 等[15]在 Qian 等的高斯过程模型基础上,引入了超球参数化对定性因子之间的关系进行建模,量化了定性因子之间的关系. Deng 等[16]进一步改进了定性因子之间的相关结构,提出了更灵活的加性高斯过程模型,并通过数值及工程案例证明了所提方法具有更好的预测性能. 然而,以上定性-定量因子混合建模方法基于高斯过程,需要对相关性矩阵进行求逆运算,要求训练样本的均布性较高,以保证求解的稳定性,且存在模型易过拟合等弊端[17]. 支持向量回归可通过将训练数据设置为支持向量及非支持向量提升模型泛化能力,并通过正则化系数控制拟合程度,而其相应的定性-定量因子混合建模方法还未见报道. 为此,本文提出一种定性-定量因子支持向量回归算法解决定性-定量因子混合的问题. 在该算法中,构建一种新型核函数描述定性-定量因子间的耦合关系,利用优化核函数超参数获取最优模型. 本文采用多个数值案例验证所提方法的有效性,核验数个经典工程问题,并通过种植牙中种植体骨应力预测的应用,检验该算法解决实际工程问题的能力.

1 算法介绍

1.1 支持向量回归

支持向量回归源于支持向量机(SVM),其通过在高维空间构建最优超平面来拟合样本,对于一组数据集$\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$,有

$$y_i=\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b;\ i=1,2,\cdots,m \tag{1}$$

其中$\boldsymbol{w}=(w_1\quad w_2\quad\cdots\quad w_m)$是权重矢量,$b$是偏置项,$\varphi(\boldsymbol{x})$为低维空间到高维空间的映射函数. 在实际的预测问题中,数据很难满足式(1)中所示回归方程. 为此,研究人员对于式(1)进行改进,设定模型输出和真实输出之间最多有$\epsilon\geqslant 0$的偏差,记ϵ为不敏感损失参数. 为了得到最大间隔的划分超平面,优化参数$\boldsymbol{w}$和b,求解如下公式:

$$\begin{aligned}&\min_{\boldsymbol{w},b}\ \frac{1}{2}\|\boldsymbol{w}\|^2\\ &\text{s.t.}\ \|y_i-(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b)\|\leqslant\epsilon\ ,\ i=1,2,\cdots,m\end{aligned} \tag{2}$$

引入正则化参数C与松弛变量ξ和$\hat{\xi}$,其优化问题可以表示为

$$\begin{aligned}&\min_{\boldsymbol{w},b,\hat{\xi}_i,\xi_i}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}(\hat{\xi}_i+\xi_i)\\ &\text{s.t.}\ y_i-(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b)\leqslant\epsilon+\xi_i\\ &\quad(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b)-y_i\leqslant\epsilon+\hat{\xi}_i\\ &\quad\xi_i\geqslant 0,\ \hat{\xi}_i\geqslant 0,\ i=1,2,\cdots,m\end{aligned} \tag{3}$$

引入拉格朗日乘子α_i和$\hat{\alpha}_i$,式(3)的对偶问题可以转化为二次规划的问题:

$$\begin{aligned}&\max_{\alpha,\hat{\alpha}}\sum_{i=1}^{m}[y_i(\hat{\alpha}_i-\alpha_i)-\epsilon(\hat{\alpha}_i+\alpha_i)]-\\ &\quad\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\alpha}_i-\alpha_i)(\hat{\alpha}_j-\alpha_j)\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j\\ &\text{s.t.}\ \sum_{i=1}^{m}\hat{\alpha}_i-\alpha_i=0\\ &\quad 0\leqslant\alpha_i,\ \hat{\alpha}_i\leqslant C,\ i=1,2,\cdots,m\end{aligned} \tag{4}$$

偏置项b为

$$b=y_i+\epsilon-\sum_{j=1}^{m}(\hat{\alpha}_j-\alpha_j)\boldsymbol{x}_j{}^{\mathrm{T}}\boldsymbol{x}_i \tag{5}$$

通过核函数$R(\boldsymbol{x},\boldsymbol{x}_i)$表示数据映射关系$\varphi(\boldsymbol{x})$,则支持向量回归的最终表达形式为

$$f(\boldsymbol{x})=\sum_{i=1}^{m}(\hat{\alpha}_i-\alpha_i)R(\boldsymbol{x},\boldsymbol{x}_i)+b \tag{6}$$

1.2 定性-定量因子支持向量回归

记输入数据中有p个定量因子$\boldsymbol{x}=(x_1\quad x_2\quad\cdots\quad x_p)^{\mathrm{T}}$和$q$个定性因子$\boldsymbol{z}=(z_1\quad z_2\quad\cdots\quad z_q)^{\mathrm{T}}$,且对于第$j$个定性因子$\boldsymbol{z}_j$有$m_j$个等级.

对于定量因子输入数据建模,支持向量回归的核函数映射关系为

$$\varphi(\cdot)=\sigma^2R(\cdot\,|\boldsymbol{\theta}) \tag{7}$$

其中σ^2是方差参数,$R(\cdot\,|\boldsymbol{\theta})$是由核函数参数向量$\boldsymbol{\theta}=(\theta_1\quad\theta_2\quad\cdots\quad\theta_p)^{\mathrm{T}}$确定的核函数. 对任意两组具有相同维度的定量因子输入量$\boldsymbol{x}_i=(x_{i1}\quad x_{i2}\quad\cdots\quad x_{ip})^{\mathrm{T}}$和$\boldsymbol{x}_j=(x_{j1}\quad x_{j2}\quad\cdots\quad x_{jp})^{\mathrm{T}}$,核函数表述如下:

$$R(\boldsymbol{x}_i,\boldsymbol{x}_j|\boldsymbol{\theta})=\exp\left\{-\sum_{k=1}^{p}\theta_k(x_{ik}-x_{jk})^2\right\} \tag{8}$$

由于式(8)的高斯核函数是基于欧几里得距离进行关系描述的,而这种关系不能描述不同等级的定性因子组合情况. 本文引入超球面分解量化定性因子之间的相关性[15]. 为得到定性因子$\boldsymbol{z}_j$的m_j个等级之间的相关性矩阵,首先构建一个严格对角元素的下三角矩阵$\boldsymbol{L}_j=(l_{r,s}^{(j)})$,$r$和$s$分别代表定性因子等级. $r=1$时,$\boldsymbol{L}_j=(l_{1,1}^{(j)})$且$l_{1,1}^{(j)}=1$;$r=2,\cdots,m_j$时,每一个$\boldsymbol{L}_j$中的行向量$(l_{r,1}^{(j)}\quad\cdots\quad l_{r,s}^{(j)}\quad\cdots\quad l_{r,r}^{(j)})$如下所示:

$$
\begin{aligned}
&l_{r,1}^{(j)}=\cos(\psi_{j,r,1})\\
&l_{r,s}^{(j)}=\sin(\psi_{j,r,1})\cdots\sin(\psi_{j,r,s-1})\cos(\psi_{j,r,s});\\
&\quad s=2,\cdots,r-1\\
&l_{r,r}^{(j)}=\sin(\psi_{j,r,1})\cdots\sin(\psi_{j,r,r-2})\sin(\psi_{j,r,r-1})
\end{aligned}
\tag{9}
$$

其中 $\psi_{j,r,s}$ 为定性因子相关性参数，约束条件为 $\psi_{j,r,s}\in(0,\pi)$. 最终得到的定性因子相关性矩阵为 $\boldsymbol{T}_j=\boldsymbol{L}_j\boldsymbol{L}_j^{\mathrm{T}}=(\tau_{r,s}^{(j)})$，是一个 $m_j\times m_j$ 的单位对角元素正定矩阵，对于每个行向量，都有 $\tau_{r,r}^{(j)}=\sum_{s=1}^{r}(l_{r,s}^{(j)})^2=1$.

结合式(8)所示定量因子高斯核函数和式(9)所示定性因子相关性矩阵，选用一种核函数组成结构，用于分析定量因子映射关系和定性因子相关性之间的组合形式[16]. 此核函数组成结构的映射关系如下所示：

$$
\varphi((\boldsymbol{x}_1,\boldsymbol{z}_1),(\boldsymbol{x}_2,\boldsymbol{z}_2))=\sum_{j=1}^{q}\sigma_j^2R(\boldsymbol{x}_1,\boldsymbol{x}_2\,|\,\boldsymbol{\theta}^{(j)})\tau_{z_{1j},z_{2j}}^{(j)}
\tag{10}
$$

其中 $\boldsymbol{x}_i=(x_{i1}\quad x_{i2}\quad\cdots\quad x_{ip})^{\mathrm{T}}$ 是定量因子，$\boldsymbol{z}_i=(z_{i1}\quad z_{i2}\quad\cdots\quad z_{iq})^{\mathrm{T}}$ 是定性因子，σ_j^2 是方差参数. 核函数 $R(\boldsymbol{x}_1,\boldsymbol{x}_2\,|\,\boldsymbol{\theta}^{(j)})$ 量化定量因子 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的相关性；$\tau_{z_{1j},z_{2j}}^{(j)}$ 是定性因子相关性矩阵 $\boldsymbol{T}_j$ 中的元素，表示第 j 个定性因子 z_{1j} 和 z_{2j} 等级之间的相关性. 由于定量因子高斯核函数矩阵具有半正定性，与参数相乘后依然为半正定矩阵，满足核函数条件，即式(10)中各个加数 $\sigma_j^2R(\boldsymbol{x}_1,\boldsymbol{x}_2\,|\,\boldsymbol{\theta}^{(j)})\tau_{z_{1j},z_{2j}}^{(j)}$ 可以视为独立的、受不同定性因子影响的定量因子核函数；而式(10)中将所有的定性因子影响核函数相加，由于核函数之和仍然是核函数[18]，从而得到最终的映射关系，使定性-定量因子混合建模得以实现.

为了使算法能更加精确地建立映射关系，本文引入灰狼优化算法（grey wolf optimizer，GWO）[19]，用于提出算法的参数寻优.

记 p 为定量因子数目，q 为定性因子数目，m_j 为第 j 个定性因子的等级数目. 由式(7)～(10)可知，核函数映射关系中一共有 $\left(q+pq+\sum_{j=1}^{q}m_j(m_j-1)\Big/2\right)$ 个待优化的参数，分别为

$$
\begin{aligned}
&\boldsymbol{\sigma}^2=(\sigma_1^2\quad\sigma_2^2\quad\cdots\quad\sigma_q^2)\\
&\boldsymbol{\theta}=(\boldsymbol{\theta}^{(1)}\quad\boldsymbol{\theta}^{(2)}\quad\cdots\quad\boldsymbol{\theta}^{(q)})\\
&\boldsymbol{\Psi}=(\boldsymbol{\Psi}^{(1)}\quad\boldsymbol{\Psi}^{(2)}\quad\cdots\quad\boldsymbol{\Psi}^{(q)})
\end{aligned}
\tag{11}
$$

$\boldsymbol{\sigma}^2$ 是式(7)中的方差参数向量；$\boldsymbol{\theta}$ 是式(7)中的核函数参数向量，对于每一个 $\boldsymbol{\theta}^{(j)}$，都有 $\boldsymbol{\theta}^{(j)}=(\theta_1^{(j)}\quad\cdots\quad\theta_p^{(j)})$；$\boldsymbol{\Psi}$ 是式(9)中的定性因子相关性参数向量，对于每一个 $\boldsymbol{\Psi}^{(j)}$，都有 $\boldsymbol{\Psi}^{(j)}=(\psi_{j,r,1}\quad\cdots\quad\psi_{j,r,r-1})$. 参数的约束条件为

$$
\begin{aligned}
&\psi_{j,r,s}\in(0,\pi)\\
&\sigma_j\geqslant 0
\end{aligned}
\tag{12}
$$

除了核函数映射关系中的参数，支持向量回归的不敏感损失参数 ϵ 和正则化参数 C 也需要优化取得合适的数值[20]. 在算法模型的建立过程中，将以上需要优化的参数设置在同一个位置向量 $\boldsymbol{P}$ 中，并根据约束条件设置合适的寻优范围，超参数优化解都由灰狼优化算法经迭代求解得到.

本文提出的定性-定量因子支持向量回归（quantitative-qualitative factors support vector regression，QQ-SVR）算法建立过程的伪代码如下：

输入：混合定性-定量因子的输入数据 $(\boldsymbol{x},\boldsymbol{z})$；输出数据 $\boldsymbol{y}$
输出：QQ-SVR 模型

1：设置 GWO 的种群规模数 A、最大迭代次数 $I_{\max}$、误差阈值 $R_{\min}$；设置迭代次数的初始值 $I\leftarrow 1$；
2：设置待优化超参数向量 $\boldsymbol{P}$ 的初始值及优化区间；
3：while $R<R_{\min}$ and $I<I_{\max}$ do
4：GWO 调整 $\boldsymbol{P}$ 中超参数位置；
5：将 $(\boldsymbol{x},\boldsymbol{z})$ 和 $\boldsymbol{P}$ 代入式(10)，得到核函数映射 φ；
6：将 φ 和 $\boldsymbol{y}$ 代入式(4)～(6)，构建 QQ-SVR 临时模型；
7：将 $(\boldsymbol{x},\boldsymbol{z})$ 代入 QQ-SVR 临时模型，计算预测结果；
8：计算预测结果与 $\boldsymbol{y}$ 之间的误差 R；
9：迭代次数 $I\leftarrow I+1$；
10：end
11：return 最终的 QQ-SVR 模型

2　数值实验

2.1　实验设计

本文通过多组测试函数验证提出的定性-定量因子支持向量回归（QQ-SVR）算法的预测性能，如表 1 所示. 实验中，采用拉丁超立方抽样方法为定量因子输入采样，定性因子输入为全析因设计，其包含所有的等级组合. 输入数据组合了定性因子和定量因子，并通过给定函数获取输出数据，生成训练样本与测试样本. 在训练样本输出数据中添加区间为 $\epsilon\sim U[-0.1,0.1]$ 的白噪声. 为检验算法对样本均布的依赖程度，设置测试函数(1)～(4)中训练样本的定量因子采样区间与测试样本不同. 与此同时，选用来自 UCI 数据库的基准测试数据集对算法进行数值实验，其输入数据同时

表1 测试函数数据集

Tab. 1 Datasets of test functions

No.	测试函数	输入数据	训练样本数	测试样本数
(1)	$y=\sin(z_1x_1)/(x_1x_2)$	训练：$x_1,x_2\sim U[0.3\pi,1.7\pi]$ 测试：$x_1,x_2\sim U[0,2\pi]$ $z_1\in\{0,0.5,1\}$	90	90
(2)	$y=\sum_{i=1}^{4} z_1^2\exp(x_i^2z_2)$	训练：$x_1,x_2,x_3,x_4\sim U[0,0.8]$ 测试：$x_1,x_2,x_3,x_4\sim U[0,1]$ $z_1,z_2\in\{1,1.2,1.5\}$	90	90
(3)	$y=\begin{cases}\sin(x_1)/x_1;\ z_1=1\\ \cos(x_1)/x_1;\ z_1=2\\ 0.1\exp(x_1)/x_1;\ z_1=3\end{cases}$	训练：$x_1\sim U[0.3\pi,1.7\pi]$ 测试：$x_1\sim U[0,2\pi]$ $z_1\in\{1,2,3\}$	120	120
(4)	$y=\begin{cases}x_1+x_2;\ z_1=1,z_2=1\\ x_1^2+x_2^2;\ z_1=1,z_2=2\\ x_1^2+x_1x_2;\ z_1=2,z_2=1\\ x_1x_2+x_2^2;\ z_1=2,z_2=2\\ x_1^3+x_2;\ z_1=3,z_2=1\\ x_1+x_2^3;\ z_1=3,z_2=2\end{cases}$	训练：$x_1,x_2\sim U[1.2,1.8]$ 测试：$x_1,x_2\sim U[1,2]$ $z_1\in\{1,2,3\}$ $z_2\in\{1,2\}$	120	120
(5)	$y=0.1[z_1(x_1x_2+x_3x_4)]^{z_2}$	$x_1,x_2,x_3,x_4\sim U[1,2]$ $z_1,z_2\in\{-1,0,1\}$	90	90
(6)	$y=z_1[z_2x_1^3+\exp(x_2)]+z_3x_1x_2$	$x_1,x_2\sim U[1,2]$ $z_1,z_2,z_3\in\{1,2,3\}$	81	81
(7)	$y=[z_1x_2+\exp(x_1)][z_2x_1+\exp(x_2)]$	$x_1,x_2\sim U[0,1]$ $z_1,z_2\in\{-1,0,1\}$	90	90
(8)	$y=\prod_{i=1}^{4}(z_ix_i+x_{5-i})$	$x_1,x_2,x_3,x_4\sim U[0,1]$ $z_1,z_2,z_3,z_4\in\{0,1,2\}$	81	81
(9)	$y=\begin{cases}10x_1^3\sin(x_1);\ z_1=1\\ 10x_1^3\tan(x_1);\ z_1=2\end{cases}$	$x_1\sim U[-\pi/4,\pi/4]$ $z_1\in\{1,2\}$	100	100
(10)	$y=\begin{cases}0.5\sin(x_1^3)+x_2;\ z_1=1\\ 0.5\exp(x_1^3)+x_2;\ z_1=2\end{cases}$	$x_1,x_2\sim U[0,2]$ $z_1\in\{1,2\}$	100	100
(11)	$y=\begin{cases}\sin(x_1^2)+x_2^2x_3x_4;\ z_1=1\\ \cos(x_1^2)+x_2x_3^2x_4;\ z_1=2\\ -\sin(x_1^2)+x_2x_3x_4^2;\ z_1=3\end{cases}$	$x_1,x_2,x_3,x_4\sim U[1,2]$ $z_1\in\{1,2,3\}$	120	120
(12)	$y=\begin{cases}10\sin(x_1)/\exp(z_2x_1);\ z_1=1\\ 10\cos(x_1)/\exp(z_2x_1);\ z_1=2\\ -10\sin(x_1)/\exp(z_2x_1);\ z_1=3\end{cases}$	$x_1\sim U[0,2\pi]$ $z_1,z_2\in\{1,2,3\}$	90	90
(13)	$y=\begin{cases}z_2x_1+\sin(x_2)\exp[\sin(x_3)];\ z_1=1\\ z_2x_1+\cos(x_2)\exp[\sin(x_3)];\ z_1=2\\ z_2x_1-\sin(x_2)\exp[\sin(x_3)];\ z_1=3\end{cases}$	$x_1,x_2,x_3\sim U[0,2\pi]$ $z_1,z_2\in\{1,2,3\}$	90	90
(14)	$y=\begin{cases}x_1\exp(x_2)+x_3x_4;\ z_1=1\\ x_1\sin(x_2)+x_3^2x_4;\ z_1=2,z_2=1\\ x_1\cos(x_2)+x_3x_4^2;\ z_1=2,z_2=2\end{cases}$	$x_1,x_2,x_3,x_4\sim U[0,\pi/2]$ $z_1,z_2\in\{1,2\}$	100	100
(15)	$y=\begin{cases}\sin(x_1^3)+x_2^2;\ z_1=1,z_2=1\\ \cos(x_1^3)-x_2^2;\ z_1=1,z_2=2\\ \sin(x_1^3)\cdot x_2^2;\ z_1=2,z_2=1\\ \cos(x_1^3)/x_2^2;\ z_1=2,z_2=2\end{cases}$	$x_1\sim U[0,2\pi]$ $x_2\sim U[1,2]$ $z_1,z_2\in\{1,2\}$	100	100
(16)	$y=\begin{cases}x_1+\sin(x_2)+x_3+\cos(x_4);\ z_1=1,z_2=1\\ x_1\sin(x_2)+x_3\cos(x_4);\ z_1=1,z_2=2\\ x_1-\sin(x_2)+x_3-\cos(x_4);\ z_1=2,z_2=1\\ \sin(x_2)/x_1+\cos(x_4)/x_3;\ z_1=2,z_2=2\\ x_3\sin(x_2)+x_1\cos(x_4);\ z_1=3,z_2=1\\ x_1\cos(x_2)+x_3\sin(x_4);\ z_1=3,z_2=2\end{cases}$	$x_1,x_3\sim U[1,2]$ $x_2,x_4\sim U[0,2\pi]$ $z_1\in\{1,2,3\}$ $z_2\in\{1,2\}$	120	120

包含定性因子和定量因子，各数据集的基本信息如表 2 所示，训练样本和测试样本皆为随机抽取.

表 2　基准测试数据集

Tab. 2　Benchmark datasets

数据集	定性特征数	定量特征数	训练样本数	测试样本数
Servo	2	2	75	92
Yacht	1	5	100	208
Auto-MPG	3	4	120	271
ENB	2	6	240	528

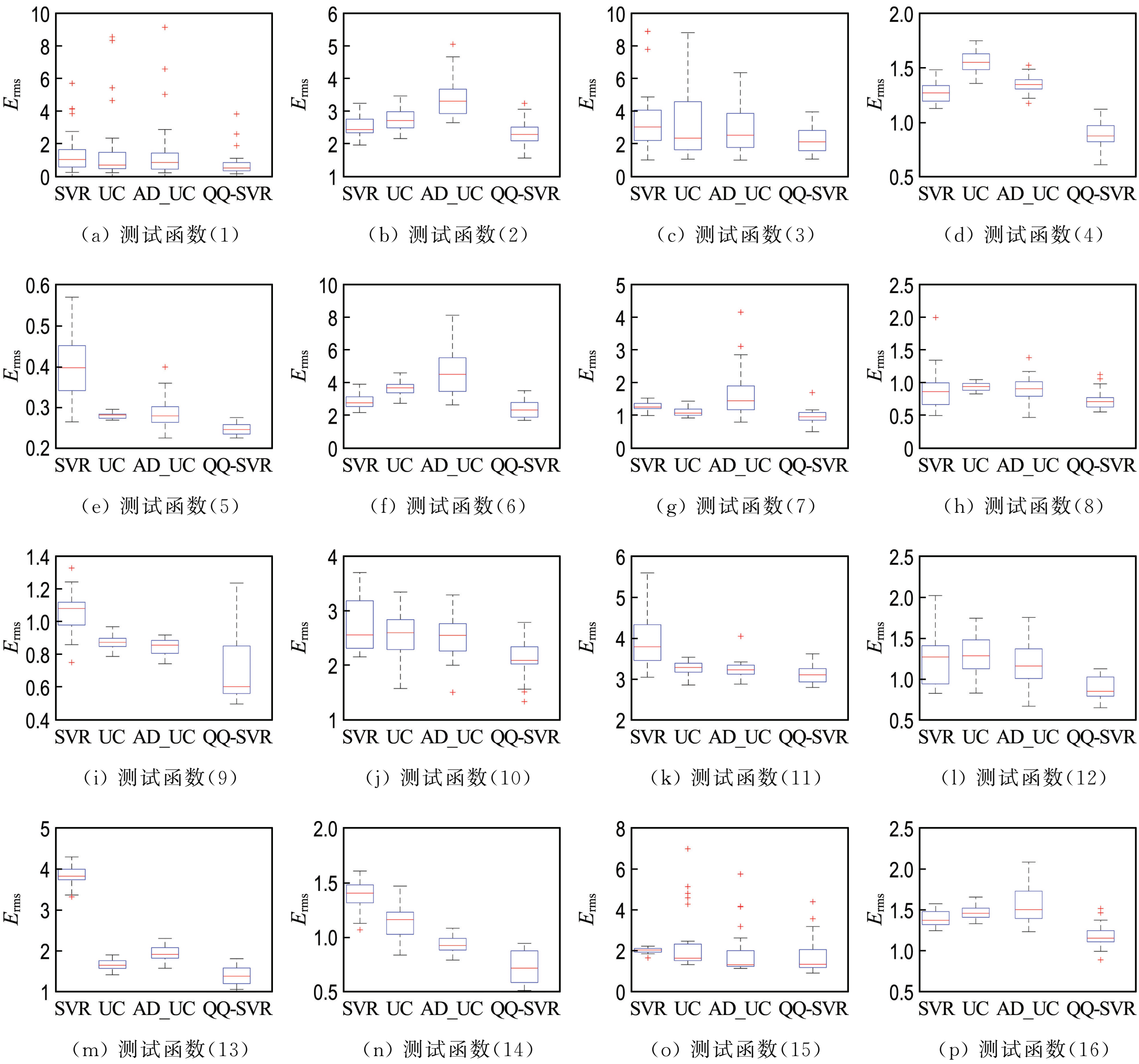
采用基于灰狼优化算法的支持向量回归(SVR)算法，文献[15-16]中的两种定性-定量因子高斯过程回归算法(UC 和 AD_UC)，以及本文提出的 QQ-SVR 算法构建预测模型. 设定 SVR 和 QQ-SVR 中用于参数寻优的 GWO 种群规模数为 20，最大迭代次数为 50，误差阈值为 0.01. 为了减少随机因素影响，每组数据集做 30 次独立分析，并以 30 次预测值的均方根误差(E_{rms})作为算法预测效果的评估指标，如下所示：

$$E_{\mathrm{rms}} = \sqrt{\sum \left\| \hat{\boldsymbol{y}}(\boldsymbol{x},\boldsymbol{z}) - \boldsymbol{y}_{\mathrm{test}}(\boldsymbol{x},\boldsymbol{z}) \right\|^2 / n} \tag{13}$$

其中$(\boldsymbol{x},\boldsymbol{z})$为测试样本的输入，$\hat{\boldsymbol{y}}(\boldsymbol{x},\boldsymbol{z})$为预测输出，$\boldsymbol{y}_{\mathrm{test}}(\boldsymbol{x},\boldsymbol{z})$为真实输出，$n$ 为测试数据样本数量. 均方根误差值越小，说明预测结果越精确.

2.2　结果与讨论

计算并统计每次实验得到的均方根误差值，获得四分位箱线图如图 1、2 所示. 可以看出，本文



(a) 测试函数(1)　(b) 测试函数(2)　(c) 测试函数(3)　(d) 测试函数(4)

(e) 测试函数(5)　(f) 测试函数(6)　(g) 测试函数(7)　(h) 测试函数(8)

(i) 测试函数(9)　(j) 测试函数(10)　(k) 测试函数(11)　(l) 测试函数(12)

(m) 测试函数(13)　(n) 测试函数(14)　(o) 测试函数(15)　(p) 测试函数(16)

图 1　均方根误差分布箱线图(一)

Fig. 1　Boxplots of RMSEs (1)

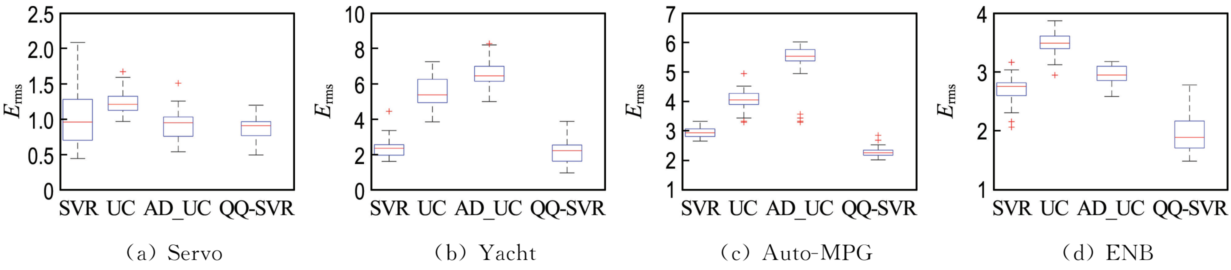


图 2 均方根误差分布箱线图(二)

Fig. 2 Boxplots of RMSEs (2)

提出的算法相比于 SVR、UC 和 AD_UC 能够提供更好的预测结果，对各组测试样本预测的均方根误差的平均值有较为显著的降低，其数值统计结果如表 3 所示.

表 3 均方根误差平均值

Tab. 3 Average values of RMSEs

数据集	均方根误差平均值			
	SVR	UC	AD_UC	QQ-SVR
(1)	1.461	1.625	1.544	0.738
(2)	2.491	2.738	3.365	2.311
(3)	3.477	3.171	2.834	2.259
(4)	1.273	1.557	1.354	0.904
(5)	0.392	0.281	0.290	0.251
(6)	2.882	3.689	4.720	2.422
(7)	1.280	1.113	1.655	0.955
(8)	0.890	0.947	0.899	0.742
(9)	1.005	0.873	0.844	0.534
(10)	2.748	2.534	2.570	2.110
(11)	3.996	3.271	3.247	3.132
(12)	1.271	1.289	1.177	0.887
(13)	3.834	1.674	1.953	1.401
(14)	1.399	1.131	0.940	0.725
(15)	2.023	2.518	2.166	1.703
(16)	1.398	1.467	1.583	1.186
Servo	1.086	1.253	0.923	0.870
Yacht	2.364	5.571	6.648	2.156
Auto-MPG	2.945	4.058	5.308	2.281
ENB	2.697	3.479	2.936	2.456

为了统计性地比较本文提出算法与其他算法预测性能，采用 t 检验对各组测试函数所得结果进行检验，零假设及备择假设如表 4 所示. 检验结果列于表 5，可以看出大部分实验结果的 t 检验得到的 p 值小于 0.05，拒绝零假设，即在多数情况下提出算法的预测误差显著低于对照算法. 提出算法在面对不同种类的定性-定量因子混合输入的数据集时，都能获得较为理想的结果.

表 4 t 检验的零假设和备择假设

Tab. 4 Null hypothesis and alternative hypothesis of t-test

零假设	备择假设
提出算法的预测误差与对照算法无明显差异	提出算法的预测误差显著低于对照算法

表 5 t 检验分析结果

Tab. 5 Results of t-test analysis

数据集	p(SVR, QQ-SVR)	p(UC, QQ-SVR)	p(AD_UC, QQ-SVR)
(1)	1.4×10^{-2}	4.2×10^{-2}	4.5×10^{-2}
(2)	4.3×10^{-2}	1.1×10^{-5}	5.5×10^{-12}
(3)	2.8×10^{-3}	2.7×10^{-2}	5.6×10^{-2}
(4)	6.7×10^{-19}	5.1×10^{-30}	1.7×10^{-23}
(5)	1.3×10^{-16}	1.7×10^{-17}	2.3×10^{-6}
(6)	3.9×10^{-4}	1.0×10^{-14}	2.7×10^{-11}
(7)	6.5×10^{-9}	2.0×10^{-3}	1.1×10^{-5}
(8)	1.7×10^{-2}	5.1×10^{-10}	8.9×10^{-4}
(9)	1.4×10^{-18}	3.7×10^{-14}	1.0×10^{-12}
(10)	1.8×10^{-7}	1.3×10^{-5}	1.9×10^{-5}
(11)	3.2×10^{-9}	1.0×10^{-2}	2.7×10^{-2}
(12)	1.7×10^{-7}	2.6×10^{-10}	3.7×10^{-6}
(13)	2.6×10^{-46}	2.1×10^{-7}	8.8×10^{-15}
(14)	5.9×10^{-26}	1.9×10^{-15}	3.8×10^{-9}
(15)	3.9×10^{-2}	4.1×10^{-2}	2.8×10^{-1}
(16)	1.9×10^{-9}	3.3×10^{-14}	3.4×10^{-10}
Servo	1.9×10^{-2}	4.2×10^{-12}	2.4×10^{-1}
Yacht	2.0×10^{-1}	2.2×10^{-24}	5.8×10^{-32}
Auto-MPG	1.3×10^{-20}	2.3×10^{-32}	3.5×10^{-28}
ENB	3.7×10^{-14}	1.5×10^{-29}	3.7×10^{-21}

通过以上数值实验的分析可以得知：与普通的支持向量回归算法相比，式(10)所示的核函数映射关系能有效地描述定性因子与定量因子之间

的耦合关系. 此外，因训练样本中存在白噪声，且测试函数(1)～(4)中训练样本的数据均布性较差，对相应数据集测试结果进行分析可以得知：与基于高斯过程回归的定性-定量因子算法(UC、AD_UC)相比，本文算法的支持向量回归结构使得它所构建的预测模型具有较好的鲁棒性和泛化性.

3　经典工程算例

将算法应用于 3 例经典工程问题，分别为弹簧设计[21]、焊接梁设计[22]和压力容器设计[23]. 这些工程问题中，输入参数同时包含定性因子与定量因子，所以适用于检验算法对于此类工程问题的建模效果. 实验中，定量因子由拉丁超立方抽样方法采样，定性因子采用包含所有组合情况的全析因设计，并通过工程函数生成训练样本和测试样本. 分别使用基于灰狼优化算法的支持向量回归(SVR)算法、两种基于高斯过程回归的定性-定量因子算法(UC 和 AD_UC)，以及本文提出的 QQ-SVR 算法建立相应的预测模型，其中算法的参数设置同 2.1 节所述.

3.1　弹簧设计

该问题的目标是预测如图 3 所示的弹簧质量，它受 3 个变量影响：导线直径(d)、线圈直径(D)以及有效圈数(N)，其中 d 和 D 是定量因子，而 N 是定性因子. 实验设计如表 6 所示. 各算法预测误差如表 7 所示，本文提出的算法预测精度最高.

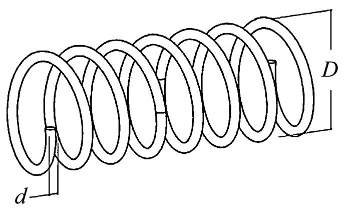


图 3　弹簧设计示意图

Fig. 3　Schematic of spring design

表 6　弹簧设计实验设计

Tab. 6　Experimental design of spring design

弹簧质量	变量范围	训练样本数	测试样本数
$y=(N+2)Dd^2$	$d \sim U[0.05, 0.20]$ $D \sim U[0.50, 1.30]$ $N \in \{2,3,4,5,6,7,8,9\}$	80	20

表 7　弹簧设计实验预测均方根误差

Tab. 7　RMSE of prediction on spring design

算法	均方根误差	算法	均方根误差
SVR	0.613	AD_UC	0.541
UC	0.585	QQ-SVR	0.492

3.2　焊接梁设计

该问题的目标是预测如图 4 所示的焊接梁成本，它受 3 个变量影响：焊缝厚度(h)、杆件附加部分长度(l)以及杆件的型号(P)，其中 h 和 l 是定量因子，P 只能被定性选取. 实验设计如表 8 所示. 各算法预测误差如表 9 所示，本文提出的算法预测精度最高.

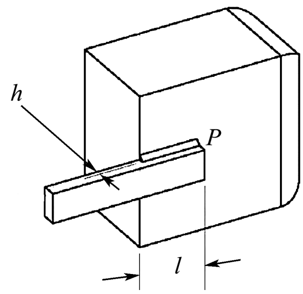


图 4　焊接梁设计示意图

Fig. 4　Schematic of welded beam design

表 8　焊接梁设计实验设计

Tab. 8　Experimental design of welded beam design

焊接梁成本	变量范围	训练样本数	测试样本数
$y=1.104\,71h^2l+0.048\,11P(14+l)$	$h \sim U[0.1, 2]$ $l \sim U[0.1, 10]$ $P \in \{1,2,3\}$	90	20

表 9　焊接梁设计实验预测均方根误差

Tab. 9　RMSE of prediction on welded beam design

算法	均方根误差	算法	均方根误差
SVR	0.350	AD_UC	0.265
UC	0.155	QQ-SVR	0.129

3.3　压力容器设计

该问题的目标是预测如图 5 所示的压力容器总成本，它受 4 个变量影响：壳体厚度(T_s)、封头厚度(T_h)、内径规格(R)以及圆柱长度规格(L)，其中 R 和 L 是定量因子，而 T_s 和 T_h 被设定为定性选取. 实验设计如表 10 所示，各算法预测误差如表 11 所示，本文提出的算法预测精度最高.

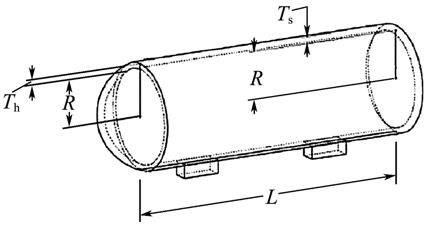


图 5 压力容器设计示意图

Fig. 5 Schematic of pressure vessel design

表 10 压力容器设计实验设计

Tab. 10 Experimental design of pressure vessel design

压力容器总成本	变量范围	训练样本数	测试样本数
$y=1.6224T_sRL+$ $1.7781T_hR^2+$ $3.1661T_s^2L+$ $19.84T_s^2R$	$R\sim U[2,9]$ $L\sim U[2,9]$ $T_s\in\{0.2,0.5,0.8\}$ $T_h\in\{0.5,0.9,1.5\}$	81	20

表 11 压力容器设计实验预测均方根误差

Tab. 11 RMSE of prediction on pressure vessel design

算法	均方根误差	算法	均方根误差
SVR	32.357	AD_UC	12.973
UC	15.092	QQ-SVR	10.952

通过此 3 个经典工程算例的实验结果可以得知：相较于其他算法，本文算法在解决受定性-定量因子混合影响的工程问题时，具有更强的建模能力，预测精度也更高. 实验结果同时也表明，本文算法适用于解决工程问题，在工程实践的设计与应用中具备一定的参考价值.

4 种植牙骨应力预测

4.1 问题描述

将本文算法进一步应用于种植牙中种植体骨应力预测以检验其工程实践意义. 如文献[24-25]所示，种植体骨应力在种植体的设计过程中是一个重要的参考指标，通常在设计过程中使用有限元分析生成数据集，再建立代理模型，通过少量采样点获取种植体结构及材料参数与骨应力在全部空间中的相关关系，实现对空间中任意一点骨应力的预测. 种植体骨应力主要受种植体结构参数和材料类型的影响，其中结构参数可按设计需求做出改变，而材料类型则相对有限. 然而现有的文献中，多数的代理模型忽略材料类型对种植体骨应力的影响，或将材料类型的选取简单等效成材料参数的数值变动，如 Roy 等[26]基于有限元分析的结果，仅针对种植牙尺寸参数建立骨应力优化的遗传算法模型；Kitagawa 等[27]基于有限元分析的结果，将不同材料的皮质骨弹性模量作为输入，建立种植牙骨应力预测模型. 这样的建模方式容易忽略材料类型变量中潜在的耦合关系，造成骨应力预测的结果不准确. 因此，有必要在建立预测模型的过程中充分考虑两类变量的混合影响. 在种植体骨应力分析中，尺寸参数可视为定量因子，材料类型可视为定性因子，在本节中，即通过本文提出的定性-定量因子支持向量回归（QQ-SVR）算法实现种植体骨应力预测.

4.2 有限元模型

构建由皮质骨、松质骨、种植体、牙冠、中央螺丝 5 部分组成的种植牙系统模型，如图 6 所示.

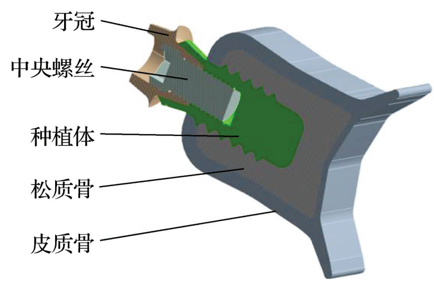


图 6 种植牙系统模型剖面图

Fig. 6 Sectional view of dental implant system model

本文选取结构参数如下：种植体长度（L_i）、螺纹长度（L_s）、螺纹螺距（P'）、等腰三角形种植体牙型底边长度（L_t）、底角角度（β），相应取值范围如表 12 所示. 种植体和牙冠可选材料共有 3 种，具体参数及相应定性因子如表 13 所示，模型其余

表 12 结构参数取值范围

Tab. 12 Value ranges of structure parameters

	L_i/mm	L_s/mm	P'/mm	L_t/mm	β/(°)
最大值	8.95	6.7	1.095	0.775	50.520
最小值	8.89	6.6	1.060	0.765	50.220

表 13 种植体和牙冠的材料参数

Tab. 13 Material parameters of implants and crowns

种植体材料参数			牙冠材料参数		
材料	弹性模量/GPa	泊松比	材料	弹性模量/GPa	泊松比
1	105	0.35	1	105	0.35
2	101	0.34	2	101	0.34
3	91	0.36	3	91	0.36

部分的材料参数如表 14 所示.

表 14　其余组成部分的材料参数

Tab. 14　Material parameters for the remaining components

组成部分	弹性模量/GPa	泊松比
皮质骨	20.0	0.30
松质骨	1.5	0.30
牙冠	220	0.30
中央螺丝	104	0.35

种植牙系统有限元网格划分如图 7 所示，其中种植体、牙冠、中央螺丝接触方式为 Frictional，摩擦因数为 0.3；种植体与骨组织接触、松质骨与皮质骨的接触类型均为 Bonded. 设置皮质骨的底面边界条件为全约束，侧面边界条件为无摩擦约束，在牙冠上加载竖直向下荷载 150 N，中央螺丝预紧力加载 150 N. 经过有限元计算后得到的种植体骨应力云图如图 8 所示.

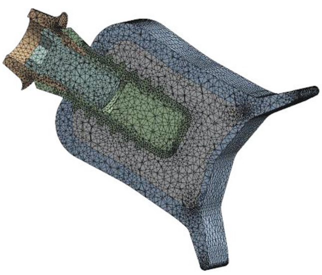


图 7　种植牙系统有限元网格划分

Fig. 7　Finite element mesh of dental implant system

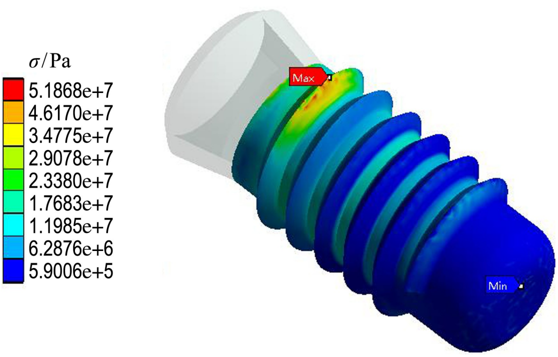


图 8　种植体骨应力云图

Fig. 8　Bone stress nephogram of implant bone

4.3　实验结果

实验中，使用拉丁超立方抽样方法对种植体模型中的结构参数进行采样，种植体和牙冠的材料采用包含所有组合情况的全析因设计. 将结构参数设定为定量因子，种植体和牙冠的材料类型编号设定为定性因子，结合成为输入数据. 将相对应的有限元模型计算得到的种植体骨应力作为输出数据. 实验一共得到 50 组输入和输出数据，随机选取 40 组作为训练样本，剩下 10 组作为测试样本. 采用基于灰狼优化算法的支持向量回归(SVR)算法、两种基于高斯过程回归的定性-定量因子算法(UC 和 AD_UC)以及本文提出的 QQ-SVR 算法分别构建种植体骨应力预测模型，算法参数设置同 2.1 节所述. 计算各模型对种植体骨应力的预测值与测试样本中真实值之间的均方根误差作为算法预测效果的评估指标. 实验结果如表 15 所示.

表 15　种植体骨应力预测均方根误差

Tab. 15　RMSE of stress prediction on implant bone

算法	均方根误差	算法	均方根误差
SVR	3.497	ADUC	6.271
UC	2.864	QQ-SVR	1.966

从表中可以看出，本文提出的 QQ-SVR 算法相比对照算法可以显著提升种植体骨应力的预测精度. 相较于 SVR、UC 和 AD_UC，本文算法对种植体骨应力预测的均方根误差分别降低了 43.8%、31.4%和 68.6%. 实验结果同时也再一次表明，此算法适用于工程实际中基于定性-定量因子混合输入的建模，具有一定的实用价值.

5　结　语

本文回顾了支持向量回归在计算机仿真和数值设计中的应用以及对定性-定量因子混合输入建模的研究现状，并针对现有算法对定性-定量因子难以有效建模的问题，提出了定性-定量因子支持向量回归算法. 在该算法中，通过一种新型核函数描述定性-定量因子耦合关系，在此基础上给出了定性-定量因子支持向量回归推导及参数优化方法，实现了定性-定量因子混合数据的有效建模. 多组计算机数值实验、经典工程算例分析以及种植体骨应力预测实验表明，所提算法相比较于现有的其他算法，对定性-定量因子混合输入的建模能力更好，能够提供更为准确的预测结果，并具有一定的工程实践能力.

参考文献：

[1]　张学工. 关于统计学习理论与支持向量机[J]. 自

动化学报，2000，**26**(1)：32-42.
ZHANG Xuegong. Introduction to statistical learning theory and support vector machines [J]. **Acta Automatica Sinica**, 2000, **26**(1): 32-42. (in Chinese)
[2] SMOLA A J, SCHOLKOPF B. A tutorial on support vector regression [J]. **Statistics and Computing**, 2004, **14**(3): 199-222.
[3] LUTE V, UPADHYAY A, SINGH K K. Computationally efficient analysis of cable-stayed bridge for GA-based optimization [J]. **Engineering Applications of Artificial Intelligence**, 2009, **22**(4/5): 750-758.
[4] EDWIN RAJA DHAS J, KUMANAN S. Evolutionary fuzzy SVR modeling of weld residual stress [J]. **Applied Soft Computing Journal**, 2016, **42**: 423-430.
[5] RAHMAN M M, MOHYALDEEN H M, NOOR M M, *et al*. Linear static response of suspension arm based on artificial neural network technique [J]. **Advanced Materials Research**, 2011, **213**: 419-426.
[6] LOSTADO R, MARTÍNEZ-DE-PISÓN F J, PERNÍA F J, *et al*. Combining regression trees and the finite element method to define stress models of highly non-linear mechanical systems [J]. **Journal of Strain Analysis for Engineering Design**, 2009, **44**(6): 491-502.
[7] 钱 骥，杨金川，李健斌，等. 基于导波奇异值向量的钢绞线应力检测方法研究[J]. 仪器仪表学报，2019，**40**(9)：27-35.
QIAN Ji, YANG Jinchuan, LI Jianbin, *et al*. Research on the stress measurement method of steel strand based on singular value vector of guided wave [J]. **Chinese Journal of Scientific Instrument**, 2019, **40**(9): 27-35. (in Chinese)
[8] 滕 军，朱焰煌，卢云军，等. 基于多输出支持向量回归机的有限元模型修正[J]. 振动与冲击，2010，**29**(3)：9-12，47.
TENG Jun, ZHU Yanhuang, LU Yunjun, *et al*. Finite element model updating based on multi-outputs support vector regression [J]. **Journal of Vibration and Shock**, 2010, **29**(3): 9-12, 47. (in Chinese)
[9] 狄勤丰，吴志浩，王文昌，等. 基于SVM的套管最大von Mises应力预测方法[J]. 石油钻探技术，2019，**47**(3)：62-67.
DI Qinfeng, WU Zhihao, WANG Wenchang, *et al*. A prediction method for determining the maximum von Mises stress in casing based on SVM [J]. **Petroleum Drilling Techniques**, 2019, **47**(3): 62-67. (in Chinese)
[10] MENDENHALL W M, SINCICH T L. **Statistics for Engineering and the Sciences** [M]. 6th ed. London: Chapman and Hall/CRC, 2016.
[11] ZHAO Yongsheng, YANG Cheng, CAI Ligang, *et al*. Surface contact stress-based nonlinear virtual material method for dynamic analysis of bolted joint of machine tool [J]. **Precision Engineering**, 2016, **43**: 230-240.
[12] SCHNEIDER R. Identification of anisotropic elastic material properties by direct mechanical simulations: estimation of process chain resource requirements [M] // Resch M, BENKERT K, WANG Xin, *et al*. eds. **High Performance Computing on Vector Systems 2010**. Berlin: Springer Berlin Heidelberg, 2010.
[13] ZHANG Yulei, NOTZ W I. Computer experiments with qualitative and quantitative variables: A review and reexamination [J]. **Quality Engineering**, 2015, **27**(1): 2-13.
[14] QIAN P Z G, WU Huaiqing, WU C F J. Gaussian process models for computer experiments with qualitative and quantitative factors [J]. **Technometrics**, 2008, **50**(3): 383-396.
[15] ZHOU Qiang, QIAN P Z G, ZHOU Shiyu. A simple approach to emulation for computer models with qualitative and quantitative factors [J]. **Technometrics**, 2011, **53**(3): 266-273.
[16] DENG X, LIN C D, LIU K W, *et al*. Additive Gaussian process for computer models with qualitative and quantitative factors [J]. **Technometrics**, 2017, **59**(3): 283-292.
[17] MOHAMMED R O, CAWLEY G C. Over-fitting in model selection with Gaussian process regression [C] // **Machine Learning and Data Mining in Pattern Recognition - 13th International Conference, MLDM 2017, Proceedings**. New York: Springer Verlag, 2017: 192-205.
[18] CRISTIANINI N, SHAWE-TAYLOR J. **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods** [M]. Cambridge: Cambridge University Press, 2000.
[19] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer [J]. **Advances in Engineering Software**, 2014, **69**: 46-61.

[20] DAI Shuyu, NIU Dongxiao, HAN Yaru. Forecasting of power grid investment in China based on support vector machine optimized by differential evolution algorithm and grey wolf optimization algorithm [J]. **Applied Sciences**, 2018, **8**(4): 636-655.

[21] COELLOC A, MONTES E M. Constraint-handling in genetic algorithms through the use of dominance-based tournament selection [J]. **Advanced Engineering Informatics**, 2002, **16**(3): 193-203.

[22] RAO S S. **Engineering Optimization: Theory and Practice** [M]. 3rd ed. New York: John Wiley & Sons, Inc., 1996.

[23] HE Qie, WANG Ling. An effective co-evolutionary particle swarm optimization for constrained engineering design problems [J]. **Engineering Applications of Artificial Intelligence**, 2007, **20**(1): 89-99.

[24] SHI Maolin, LI Hongyou, LIU Xiaomei. Multidisciplinary design optimization of dental implant based on finite element method and surrogate models [J]. **Journal of Mechanical Science and Technology**, 2017, **31**(10): 5067-5073.

[25] LI Hongyou, SHI Maolin, LIU Xiaomei, *et al*. Uncertainty optimization of dental implant based on finite element method, global sensitivity analysis and support vector regression [J]. **Proceedings of the Institution of Mechanical Engineers, Part H-Journal of Engineering in Medicine**, 2019, **233**(2): 232-243.

[26] ROY S, DEY S, KHUTIA N, *et al*. Design of patient specific dental implant using FE analysis and computational intelligence techniques [J]. **Applied Soft Computing**, 2018, **65**: 272-279.

[27] KITAGAWA T, TANIMOTO Y, NEMOTO K, *et al*. Influence of cortical bone quality on stress distribution in bone around dental implant [J]. **Dental Materials Journal**, 2005, **24**(2): 219-224.

A qualitative-quantitative factors mixed data modeling method based on support vector regression

FENG Xiangyu, SHI Maolin*, MA Yue, SONG Xueguan, SUN Wei, TIAN Teng

(School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: In the process of scientific research and engineering practice, it is very common to conduct engineering experiments and numerical simulation in which the input parameters have both qualitative and quantitative factors. To achieve effective modeling of such kind of data, a method for modeling qualitative and quantitative factors based on support vector regression (SVR) is proposed for qualitative-quantitative factors analysis in engineering experiments and numerical simulation. It quantizes the correlation between qualitative factors by using the hypersphere decomposition, and describes the correlation between qualitative factors and quantitative factors by constructing a special kernel function. A support vector regression algorithm for qualitative-quantitative factors is constructed for mixed data modeling and prediction of qualitative and quantitative data. Numerical experiments and classical engineering problems show that the proposed algorithm can provide better prediction results compared with ordinary support vector regression algorithm and qualitative-quantitative factor algorithms based on Gaussian process regression. Taking bone stress analysis of implant as an example, the type of implant material is considered as qualitative factors and the structural parameters as quantitative factors. Experimental results show that the proposed algorithm can significantly improve the accuracy of bone stress prediction and provide a model basis for implant design optimization, which verifies the engineering rationality of the proposed algorithm.

Key words: mixed data; kernel function; support vector regression (SVR); computer experiment