集成数据挖掘知识的可解释最优超球体支持向量机

陆思洁[1], 范 崠[1], 渐 令[2], 郜传厚[1†]

(1. 浙江大学 数学科学学院, 浙江 杭州 310027; 2. 中国石油大学(华东) 经济管理学院, 山东 青岛 266580)

摘要: 最优超球体支持向量机(SSLM)是一种典型的黑箱模型, 其运行模式不需要考察被研究对象的内部结构和机理, 仅利用对象的输入输出数据即能达到认识其功能和作用机制, 因此具有响应快、实时性强等优点, 但也因此缺乏可解释性和透明性. 鉴于此, 本文研究从SSLM黑箱模型的输入端加入先验知识的方法, 增强其可解释性. 本文开发了基于数据的非线性圆形知识挖掘算法以及知识的离散化算法, 离散后的数据点不仅包含产生知识的原始数据点, 还增加了新的数据点. 通过将所挖掘的圆形知识以不等式约束的形式集成至SSLM模型, 构造了可解释的SSLM模型(i-SSLM). 该模型在训练时要确保知识约束的数据点分类正确, 因此对模型结果有一定程度的预知, 表明模型具有可解释性; 同时, 又由于知识的离散化增加了新的数据信息, 因此, 模型能具有更高的精度. i-SSLM模型的有效性在10组公共样本集和2组实际高炉数据集上得到了验证.

关键词: 黑箱模型; 可解释性; 最优超球体支持向量机; 先验知识; 不平衡数据

引用格式: 陆思洁, 范崠, 渐令, 等. 集成数据挖掘知识的可解释最优超球体支持向量机. 控制理论与应用, 2024, 41(3): 375 – 384

DOI: 10.7641/CTA.2023.20832

Interpretable small sphere and large margin support vector machine with integrated data mining knowledge

LU Si-jie[1], FAN Di[1], JIAN Ling[2], GAO Chuan-hou[1†]

(1. School of Mathematical Sciences, Zhejiang University, Hangzhou Zhejiang 310027, China;
2. School of Economics and Management, China University of Petroleum (East China), Qingdao Shandong 266580, China)

Abstract: Small sphere and large margin support vector machine (SSLM) is a typical black box model, which works in no need of understanding the internal structure and mechanism of the object to be studied while only utilizes the input and output data for the purpose of knowing its function and interaction relation. Hence, the SSLM has the advantages of fast response and strong real-time performance, but accordingly lacks interpretability and transparency. In view of this, this paper examines ways to add prior knowledge into the input-port of the SSLM black box model to enhance its interpretability. We developed a nonlinear circular knowledge mining algorithm based on data as well as a discretization algorithm for knowledge, and the discrete data points contain not only the original data points that generated the knowledge, but also add new data points. By integrating the mined circular knowledge into the SSLM model in the form of inequality constraints, we construct an interpretable SSLM model (i-SSLM). When the model is trained, it is necessary to ensure that the data point classification of the knowledge constraint is correct, so there is a certain degree of prediction of the model results, indicating that the model is interpretable. At the same time, due to the discretization of knowledge to add new data information, the model can have higher accuracy. The validity of the i-SSLM model was verified on 10 sets of common sample sets and 2 sets of actual blast furnace datasets.

Key words: black box model; interpretability; small sphere and large margin support vector machine; prior knowledge; unbalanced data

Citation: LU Sijie, FAN Di, JIAN Ling, et al. Interpretable small sphere and large margin support vector machine with integrated data mining knowledge. *Control Theory & Applications*, 2024, 41(3): 375 – 384

收稿日期: 2022−09−22; 录用日期: 2023−04−24.

†通信作者. E-mail: gaochou@zju.edu.cn; Tel.: +86 571-87952431.

本文责任编委: 周平.

国家自然科学基金项目(12320101001, 12071428, 62111530247), 浙江省自然科学基金重点项目(LZ20A010002)资助.

Supported by the National Natural Science Foundation of China (12320101001, 12071428, 62111530247) and the National Natural Science Foundation of Zhejiang Province (LZ20A010002).

1 引言

传统的黑箱建模方法, 如支持向量机(support vector machine, SVM)、神经网络等, 由于不涉及被研究对象的内部结构和作用机制, 而仅利用其输入输出关系即可从整体上把握其行为, 因此受到广泛关注并常被用来处理一些复杂系统, 如人脑[1]、黑洞[2]、工业过程[3]等. 可以说, 黑箱建模提供了一种认识复杂对象的有效途径. 但近年其发展也受到诸多挑战, 其中最重要的一条是黑箱模型的运行方式缺乏可解释性及透明性, 不能把模型结果直接转换为能被理解的知识; 同时, 也不能利用问题本身具有的先验知识改进模型性能, 因此, 模型结果的说服力和对问题的实际指导价值都减弱, 这在很大程度上限制了黑箱模型更为广泛、深入的应用. 为此, 研究黑箱模型的可解释性就变得十分重要[4–7], 特别是在需要对作出的决定给出明确解释的领域, 如信用风险分析、医疗诊断等.

黑箱模型的可解释性在数学上并没有严格、统一的定义, 通常理解为人们对决策的理解程度或对模型结果的预知程度. 因此, 增强黑箱模型的可解释性有以下两个途径: 1) 从黑箱模型的输出端提取规则, 将系统中隐含的知识以一种易于理解的方式表达出来[8–12]; 2) 在黑箱模型的输入端集成先验知识[13–17]. 其中先验知识是指目标系统建模前就已知的任何相关信息(包括领域知识、专家经验、数据等), 并且在有限数据条件下, 加入先验知识是提高黑箱模型泛化性能的唯一手段[18]. 本文主要关注第2种方式, 即在黑箱模型的输入端集成先验知识以增强其可解释性.

先验知识的种类繁多, 其集成黑箱模型的方式也因此而不同. 以SVM黑箱建模技术为例, 其先验知识的集成方法可分为3类: 1) 结构改进. 该方法主要通过合理地设计核函数的形式来实现先验知识集成, 如引入变换不变性构造新核函数[19]、定义新的距离代替欧氏距离构造新核函数[20]、利用Haar积分构造新核函数[21]以及利用先验知识直接构造核函数[22–23]等; 2) 算法改进. 该方法通过SVM本身算法的设计来集成先验知识, 如通过在SVM优化问题中增加相应的线性或非线性不等式约束来实现线性或非线性先验知识的集成[24–25]、利用半定规划算法实现变换不变性先验知识的集成[26]等; 3) 数据样本. 该方法主要通过对样本数量或样本重要程度的控制来集成先验知识, 如利用先验知识产生更多训练样本的虚拟样本法[27]、通过设置正则化权重系数融合样本分布不平衡[28]、样本准确度[29]和引入新的数据集Universum[30–31]等先验知识的权重约束法.

从数据中挖掘先验知识并进一步集成至黑箱模型已展现出较好的透明化效果. 笔者之前的工作[32–33]实现了从数据中挖掘二维线性先验知识, 并以线性不等式约束的形式集成至软间隔SVM模型. 通过将集成知识后的模型改造成标准的二次规划问题, 得到了部分可解释的软间隔支持向量机模型, 并在一些公共样本集和实际高炉数据集上展现了可解释性增强效果. 但因挖掘的知识为线性先验知识, 且在形成不等式约束时仍仅利用产生知识的数据点(即本已存在的数据点), 对于构造的模型并没有增加新的数据信息, 导致透明化后的模型精度提升并不明显. 特别是对比例偏低的不平衡样本点, 新模型几乎没有效果; 并且, 模型的评价指标仅采用预测精度, 未能合理体现处理不平衡数据集的模型的真实性能. 为此, 本文以最小体积最大间隔支持向量机(small sphere and large margin, SSLM, 下文称为最优超球体支持向量机)[34]为基准模型(其工作原理是通过构建超球体执行分类任务, 并且常用来处理不平衡数据集), 开发从数据中挖掘非线性圆形知识的算法. 为更好的处理数据的不平衡性, 进一步开发圆形知识所包含区域的数据点离散化算法, 使得知识所约束的数据点不仅包含产生知识的原有数据点, 还产生新的数据点. 最后, 将挖掘的非线性圆形知识集成到SSLM黑箱模型, 构建可解释的最优超球体支持向量机模型, 记为i-SSLM. 这里模型的效果通过12组数据集(包括10组公共样本集和2组实际高炉数据集)以及选择对比传统的4种模型C-SVM[35], 支持向量数据描述(support vector data descriptio, SVDD)[36], SSLM[34] 和最大间隔双球支持向量机(maximum margin of twin spheres support vector machine, MMTSSVM)[37]加以展示. 相比较于这些黑箱模型, i-SSLM模型本身具有能有效处理不平衡数据的优势; 其次, 它通过融入非线性先验知识可对输出结果进行解释; 最后, 它通过离散化知识所框区域, 增加新的样本信息, 有望提高模型的分类性能.

论文的结构如下: 第2节介绍了SSLM模型的工作原理及逻辑知识如何转化为不等式; 第3节给出一种从数据中挖掘圆形先验知识的算法并将挖掘的知识以不等式约束的形式集成至SSLM模型, 构造具有可解释性的i-SSLM模型; 第4节通过10个公共样本集和2个高炉实际样本集验证i-SSLM模型的有效性, 并与传统的C-SVM, SVDD, SSLM, MMSSVM这4种模型进行对比; 第5节总结全文并对将来工作进行展望.

符号说明: 文中小写粗体字母$\boldsymbol{a}$表示列向量, 大写粗体字母$\boldsymbol{A}$表示矩阵, 上标$\boldsymbol{A}^{\mathrm{T}}$表示矩阵$\boldsymbol{A}$的转置; $\mathbb{R}^m$表示m维实数空间, $\mathbb{R}^m_+$ 表示m维正实数空间; 对任意$\boldsymbol{x}\in\mathbb{R}^m$, $\|\boldsymbol{x}\|$表示2–范数, $<\boldsymbol{x},\boldsymbol{x}>$表示内积; $\mathbf{1}$和$\mathbf{0}$分别表示分量全为1和0的列向量.

2 预备知识

本节将简单介绍最优超球体支持向量机的工作原理和逻辑知识的表达及转换.

2.1　最优超球体支持向量机

最优超球体支持向量机(SSLM)常用来处理不平衡数据集的分类问题, 针对数据集$\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$(假设$\boldsymbol{x}_i \in \mathbb{R}^m$, $y_i \in \{+1, -1\}$, 数据集中正类样本个数n^+远大于负类样本个数n^-), 其工作原理是通过构建一个尽可能多地包围正类样本点的超球来完成分类任务. 若样本点落在超球内, 则标识为正类; 若样本点落在超球外, 则标识为负类. 同时, 为了实现最大程度的分离, 要求球面到负类样本点的距离应尽可能大. 数学上, SSLM[34]可表示为

$$\begin{aligned}&\min_{R,\boldsymbol{c},\rho,\boldsymbol{\xi},\boldsymbol{\eta}} R^2-\nu\rho^2+\frac{1}{\nu_1 n^+}\sum_{i=1}^{n^+}\xi_i+\frac{1}{\nu_2 n^-}\sum_{j=1}^{n^-}\eta_j,\\ &\text{s.t. } \|\phi(\boldsymbol{x}_i^+)-\boldsymbol{c}\|^2\leqslant R^2+\xi_i,\ \xi_i\geqslant 0,\\ &\quad \|\phi(\boldsymbol{x}_j^-)-\boldsymbol{c}\|^2\geqslant R^2+\rho^2-\eta_j,\ \eta_j\geqslant 0,\\ &\quad 1\leqslant i\leqslant n^+,\ 1\leqslant j\leqslant n^-,\end{aligned}\tag{1}$$

其中: R和$\boldsymbol{c}$分别为超球的半径和球心; $\rho\in\mathbb{R}$, $\rho^2\geqslant 0$表示超球表面与负类点的间隔; $\boldsymbol{x}_i^+$和$\boldsymbol{x}_j^-$分别为正类样本点和负类样本点; $\phi:\mathbb{R}^m\to\mathcal{F}$是一个从$\mathbb{R}^m$到无穷维空间$\mathcal{F}$的高维映射; ξ_i和η_j为松弛变量; ν, ν_1, ν_2是3个正常数.

通过引入拉格朗日乘子$\boldsymbol{\alpha}=[\alpha_1\ \ \alpha_2\ \ \cdots\ \ \alpha_n]^{\mathrm{T}}$并进行对偶变换, 可求解得模型(1)的决策函数为

$$\begin{aligned}&f(\boldsymbol{x})=\operatorname{sgn}(R^2-\|\phi(\boldsymbol{x})-\boldsymbol{c}\|^2)=\\ &\operatorname{sgn}(R^2-\langle\boldsymbol{c},\boldsymbol{c}\rangle-K(\boldsymbol{x},\boldsymbol{x})+2\sum_{i=1}^{n}\alpha_i y_i K(\boldsymbol{x},\boldsymbol{x}_i)),\end{aligned}\tag{2}$$

其中$K(\boldsymbol{x},\boldsymbol{x})=\langle\phi(\boldsymbol{x}),\phi(\boldsymbol{x})\rangle$为核函数. 当$f(\boldsymbol{x})\geqslant 0$, 则$y=+1$; 当$f(\boldsymbol{x})<0$, 则$y=-1$.

2.2　知识的表达及转换

知识的定义有许多种, 而关于对象的逻辑推理式是其中一种重要的表现形式. 以上述二分类问题为例, 则其逻辑知识可以表达为

$$\boldsymbol{g}(\boldsymbol{x})\leqslant 0\Rightarrow\boldsymbol{x}\in\mathcal{X}^+\text{或}f(\boldsymbol{x})\geqslant 0,\ \forall\boldsymbol{x}\in\Gamma,\tag{3}$$

式中: $\boldsymbol{g}:\Gamma\to\mathbb{R}^k$是定义在$\Gamma$上的$k$维函数, $\Gamma\subseteq\mathbb{R}^m$; $\mathcal{X}^+$表示正类样本集合. 该式的物理意义是集合$\{\boldsymbol{x}|\boldsymbol{g}(\boldsymbol{x})\leqslant 0\}$中所有点都属于正类, 类似可表达负类逻辑知识. 注意这种以逻辑形式给出的知识并不能直接融入SSLM模型, 需将其转化为方便集成的模式.

数学上, 式(3)中的逻辑表达式与下述方程组等价[38]:

$$\left.\begin{aligned}g(\boldsymbol{x})\leqslant 0,\\ f(\boldsymbol{x})<0,\end{aligned}\right\}\text{无解},\ \forall\boldsymbol{x}\in\Gamma,\tag{4}$$

进一步, 他们证明了若式(3)或式(4)成立, 则存在$\boldsymbol{u}\in\mathbb{R}_+^k$, 使得

$$f(\boldsymbol{x})+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})\geqslant 0,\ \forall\boldsymbol{x}\in\Gamma,\tag{5}$$

显然, 式(5)可作为约束直接集成到SSLM模型中.

3　可解释的最优超球体支持向量机

在第3节中, 将开发算法从数据中挖掘圆形非线性知识, 并将其集成到SSLM模型中以获取可解释性.

3.1　数据型非线性知识挖掘算法

从数据中挖掘知识, 关键是给出式(5)中$\boldsymbol{g}(\boldsymbol{x})$的具体表达式, 这在特征维数较高、函数关系较为复杂时通常十分困难. Chen等人[32–33]提出一种基于数据的双变量的线性先验知识挖掘方法, 通过将特征空间投影到二维空间, 并在所有二维特征空间中挑选能最大分离正类(或负类)样本的线性先验知识; 同时挖掘的线性先验知识以不等式约束的形式集成到黑箱SVM模型中, 创建一个能同步优化挖掘知识的正确性和黑箱模型的优化问题. 受Chen等[32–33]启发, 本文只考虑由系统两个特征变量所生成的知识, 即$g:\Gamma\subseteq\mathbb{R}^2\to\mathbb{R}$, 意味着把样本点从输入空间$\mathbb{R}^m$映射到一个二维子空间, 用$X^{(i)}OX^{(j)}(i,j=1,\cdots,m)$来表示. 同时, 考虑到SSLM的工作特点(构造最优超球体进行分类)和样本非平衡性(正类样本远多于负类样本), 本文寻找二维平面$X^{(i)}OX^{(j)}$上的圆形正类知识, 即

$$\begin{aligned}\mathcal{C}^+:g(x^{(i)},x^{(j)})=&\|(x^{(i)},x^{(j)})-(\tan\theta_1,\tan\theta_2)\|^2-\\ &r^2\leqslant 0\Rightarrow\boldsymbol{x}\in\mathcal{X}^+,\end{aligned}\tag{6}$$

式中: $(\tan\theta_1,\tan\theta_2)$为圆心, r为圆半径. 这里圆心的设置是通过正切函数实现, θ_1和θ_2在区间$(-\frac{\pi}{2},\frac{\pi}{2})$中均匀选取10个点进行交叉组合设置. 具体算法如算法1(见表1)所示.

步骤7的优化问题保证正类样本点尽可能多的在圆内, 约束表示所有的负类样本点都要在圆外, 即在二维子空间中, 正类样本点尽可能多的分类正确, 而负类样本点全部分类正确. 实际求解时, 每次循环均固定两个特征x_i与x_j及θ_1和θ_2的取值, 在约束条件成立情况下, 计算目标函数值, 即正类样本点分类正确的个数, 循环结束时目标函数最大值即为所求. 在最后的输出结果中, 除了得到圆形知识之外, 圆内所包含的正类样本点也同时得到.

注 1　算法1挖掘的知识将以约束的形式集成到SSLM模型, 因其能确保模型在圆内的正类样本点分类正确, 即$\mathcal{I}_1$内的样本点, 所以集成这类知识后的SSLM具有一定的可解释性. 同时, 模型的精度也有望提高, 因此确保了一部分样本分类正确.

例 1　本文将算法1应用于UCI数据库里的Liver Disorders数据集[1]. 该数据集共含有6个特征变量, 345个样本点, 其中正类样本点200个, 负类样本点145个. 为构建不平衡数据集, 本文随机选取70%的正类点,

[1]https://archive-beta.ics.uci.edu/ml/datasets/liver+disorders

即140个正类点, 和16个负类点, 使得正类点与负类点比例大约为$9:1$. 应用算法1, 最后可得在$X^{(2)}OX^{(3)}$二维子空间, 圆内所包含的正类点个数最多, 为45个正类点. 因此, 选择

$$g(x^{(2)},x^{(3)})=\\ \|(x^{(2)},x^{(3)})-(-1.4586,-3.2213)\|^2-13.9616, \tag{7}$$

用于后面表达该数据集挖掘的非线性知识.

表 1 算法1: 数据挖掘圆形非线性知识

Table 1 Algorithm 1: Circular nonlinear knowledge mined from data

Input: $\{(\boldsymbol{x}_k,y_k)\}_{k=1}^n$, $\boldsymbol{x}=[x^{(1)}\ \ \cdots\ \ x^{(m)}]^{\mathrm{T}}\in\mathbb{R}^m$, $y_k\in\{+1,-1\}$;
Output: 挖掘的正类知识;
1 数据归一化;
2 记$\boldsymbol{\tau}=(\tau_1,\tau_2,\cdots,\tau_{10})$为区间$(-\frac{\pi}{2},\frac{\pi}{2})$中均匀选取的10个点;
3 **for** $i=1,2,\cdots,m-1$ **do**
4 **for** $j=i+1,i+2,\cdots,m$ **do**
5 **for** $\theta_1=\tau_1,\tau_2,\cdots,\tau_{10}$ **do**
6 **for** $\theta_2=\tau_1,\tau_2,\cdots,\tau_{10}$ **do**
7 求解如下优化问题:

$$\max_{\theta_1,\theta_2,r}\ \mathrm{Card}(\{\boldsymbol{x}\in\mathcal{X}^+|\,\|(x^{(i)},x^{(j)})-(\tan\theta_1,\tan\theta_2)\|^2-r^2\leqslant 0\}),$$
$$\text{s.t.}\ \ \|(x^{(i)},x^{(j)})-(\tan\theta_1,\tan\theta_2)\|^2-r^2>0,\ \ \forall\boldsymbol{x}\in\mathcal{X}^-,$$

其中: Card(·) 表示集合中元素的个数, $r^2=\min(\|(x^{(i)},x^{(j)})-(\tan\theta_1,\tan\theta_2)\|^2)-\epsilon$, $\boldsymbol{x}\in\mathcal{X}^-$, ϵ为一个趋近于0的数;
8 把最优解保存到$\Omega_{(i,j;\hat{\theta}_1,\hat{\theta}_2,\hat{r})}$中;
9 **end for**
10 **end for**
11 **end for**
12 **end for**
13 $(\hat{i},\hat{j};\hat{\theta}_1,\hat{\theta}_2,\hat{r})=\arg\max_{i,j}\Omega_{(i,j;\hat{\theta}_1,\hat{\theta}_2,\hat{r})}$;
14 从样本点中挖掘到的正类先验知识为

$$\mathcal{I}_1=\{\boldsymbol{x}_p|\,\|(x_p^{(\hat{i})},x_p^{(\hat{j})})-(\tan\hat{\theta}_1,\tan\hat{\theta}_2)\|^2-\hat{r}^2\leqslant 0\Rightarrow \boldsymbol{x}_p\in\mathcal{X}^+\}.$$

注 2 式(7)的知识是由140个正类样本点所生成, 但在圆形知识形成后, 由式(5)知, 圆形图所包含的整个区域都属于正类知识, 将其直接集成至SSLM模型将会得到一个半无穷规划问题, 难以求解. 为了尽可能利用区域知识, 同时也为了方便集成SSLM模型, 本文对区域里的知识离散化来获取更多的正类样本点, 算法2(见表2)即为实现这一目的而开发, 其主要思想是将$\mathcal{I}_1$ 中的l个点作为网格点形成网格, 并在圆形区域内随机选取除$\mathcal{I}_1$以外的其它$2l$个网格交点作为虚拟正类样本点, 最终的知识点集合为$\mathcal{I}=\mathcal{I}_1\cup\mathcal{I}_2$. 当然, 这些新增加的样本信息作为约束将会增加模型的复杂度, 进而可能对模型训练不利, 笔者后面将通过引入松弛因子来强化它们的正确性.

表 2 算法2: 圆形知识离散化

Table 2 Algorithm 2: Discretization of circular knowledge

Input: $\mathcal{I}_1=\{\boldsymbol{x}_p\}_{p=1}^l$, $\boldsymbol{x}_p=[x_p^{(1)}\ \ x_p^{(2)}\ \ \cdots\ \ x_p^{(m)}]^{\mathrm{T}}$, $l=\mathrm{Card}(\mathcal{I}_1)$, $\mathcal{I}_2=\varnothing$;
Output: 生成的虚拟知识;
1 令$\mathrm{num}=1$, $\boldsymbol{z}=\boldsymbol{0}\in\mathbb{R}^m$;
2 **while** $\mathrm{num}\leqslant 2l$ **do**
3 随机生成m个$1\sim l$间的整数, 记为$\{h_1,h_2,\cdots,h_m\}$;
4 $\boldsymbol{z}_{\mathrm{num}}=(x_{h_1}^{(1)},x_{h_2}^{(2)},\cdots,x_{h_m}^{(m)})$, 其中$x_{h_m}^{(m)}$表示$\mathcal{I}_1$中样本点所组成矩阵的第$h_m$行第$m$列的数;
5 **if** $\boldsymbol{z}_{\mathrm{num}}\in\mathcal{I}_1\cup\mathcal{I}_2$ **then**
6 返回步骤3;
7 **else**
8 $\mathcal{I}_2=\mathcal{I}_2\cup\{\boldsymbol{z}_{\mathrm{num}}\}$;
9 **end if**
10 $\mathrm{num}=\mathrm{num}+1$.
11 **end while**

基于式(3)–(5), 离散化后的圆形知识可以表示为

$$f(\boldsymbol{x}_p)+ug(x_p^{(i)},x_p^{(j)})\geqslant 0,\ p=1,\cdots,3l,\ \forall\boldsymbol{x}_p\in\mathcal{I}, \tag{8}$$

其中: $u>0$, $g(x_p^{(i)},x_p^{(j)})$见式(7).

3.2 圆形知识集成的最优超球体支持向量机

将形如式(8)的知识集成到SSLM模型可得

$$\begin{aligned}\min_{R,\boldsymbol{c},\rho,u,\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{z}}\ & R^2-\nu\rho^2+\frac{1}{\nu_1 n^+}\sum_{i=1}^{n^+}\xi_i+\frac{1}{\nu_2 n^-}\sum_{j=1}^{n^-}\eta_j+\\ &\delta\sum_{p=1}^{t}z_p,\\ \text{s.t.}\ \ & \|\phi(\boldsymbol{x}_i^+)-\boldsymbol{c}\|^2\leqslant R^2+\xi_i,\ \xi_i\geqslant 0,\\ & \|\phi(\boldsymbol{x}_j^-)-\boldsymbol{c}\|^2\geqslant R^2+\rho^2-\eta_j,\ \eta_j\geqslant 0,\\ & f(\boldsymbol{x}_p)+ug(x_p^{(t)},x_p^{(q)})+z_p\geqslant 0,\\ & 1\leqslant i\leqslant n^+,\ 1\leqslant j\leqslant n^-,\\ & u>0,\ z_p\geqslant 0,\ 1\leqslant p\leqslant 3l,\ \boldsymbol{x}_p\in\mathcal{I},\end{aligned} \tag{9}$$

其中: δ与ν,ν_1,ν_2意义相同, 表权重参数; z_p为松弛变量, 测量圆形知识的偏差; $x_p^{(t)}$, $x_p^{(q)}$为$\boldsymbol{x}_p$在$X^{(t)}OX^{(q)}$平面上的投影分量. 为区别SSLM, 称模型(9)为可解释的最优超球体支持向量机, 记为i-SSLM.

通过引入拉格朗日乘子并做对偶变换, 再进一步利用KKT(Karush-Kuhn-Tucker)条件, 可得i-SSLM的决策函数为

$$f(\boldsymbol{x})=\mathrm{sgn}(R^2-\|\phi(\boldsymbol{x})-\boldsymbol{c}\|^2)=$$
$$\mathrm{sgn}(R^2-\langle\boldsymbol{c},\boldsymbol{c}\rangle-K(\boldsymbol{x},\boldsymbol{x})+2\sum_{i=1}^{n}\alpha_i^1y_iK(\boldsymbol{x},\boldsymbol{x}_i)+$$
$$2\sum_{p=1}^{3l}\alpha_p^2K(\boldsymbol{x},\boldsymbol{x}_p)). \tag{10}$$

其中: $R^2=\dfrac{1}{n_1}P_1$,

$$\begin{aligned}\langle\boldsymbol{c},\boldsymbol{c}\rangle=&\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i^1\alpha_j^1y_iy_jK(\boldsymbol{x}_i,\boldsymbol{x}_j)+\\&\sum_{p_1=1}^{3l}\sum_{p_2=1}^{3l}\alpha_{p_1}^2\alpha_{p_2}^2k(\boldsymbol{x}_{p_1},\boldsymbol{x}_{p_2})+\\&2\sum_{i=1}^{n}\sum_{p=1}^{3l}\alpha_i^1\alpha_p^2y_iK(\boldsymbol{x}_i,\boldsymbol{x}_p).\end{aligned} \tag{11}$$

这里

$$\begin{cases}n_1=|\mathcal{S}_1|,\\P_1=\sum\limits_{\boldsymbol{x}_i\in\mathcal{S}_1}(K(\boldsymbol{x}_i,\boldsymbol{x}_i)-\\\qquad 2\sum\limits_{k=1}^{n}\alpha_ky_kK(\boldsymbol{x}_i,\boldsymbol{x}_k)+\langle\boldsymbol{c},\boldsymbol{c}\rangle),\\\mathcal{S}_1=\{\boldsymbol{x}_i|0<\alpha_i<\dfrac{1}{\nu_1n^+},\ 1\leqslant i\leqslant n^+\}.\end{cases}$$

注 3　本质上, i-SSLM模型(9)是通过集成先验知识和增加训练样本来增强黑箱SSLM模型的可解释性, 但这并不是对现有知识集成[32–33]及产生虚拟样本方法[27]的组合. i-SSLM 综合利用了SSLM模型能有效处理不平衡数据、集成的非线性圆形先验知识与SSLM模型类型相一致、增加的额外离散知识点提供了新信息等优势来提高分类性能. 这里产生知识的方法以及增加离散样本点的方法都与上述文献不同, 文献[32–33]产生的是线性先验知识, 而文献[27]离散样本的方法是通过使用形态学运算符人为地细化和粗化线条来为每个示例生成两个额外的向量.

3.3　公共数据集实验

本文首先在10个公共数据集上进行实验, 这些数据集均有相应的实际背景, 它们的基本信息见表3, 其中最后一列“Example size”是按照“#Example×#Feature”计算而得, 用以衡量数据集的规模. 对每个数据集, 从正类样本中随机选取70%样本, 同时从负类样本中按照正负比例约为9 : 1的数量随机选取样本构成训练集, 剩下的正类、负类样本作为测试集. 很显然, 训练集具有不平衡性. 按照 i-SSLM 建模步骤, 首先利用算法1在训练集上生成圆形非线性知识(区域), 进一步利用算法2将圆形区域知识离散化并集成到i-SSLM模型形成式(9), 最后通过训练集参数寻优和测试集验证可得模型的测试精度.

表 3　10组公共数据集基本信息

Table 3　10 sets of public dataset basic information

数据集	#Example	#Feature	#Positive	#Negative	Example size
Arrhythmia	420	278	183	237	116760
Australian	690	14	307	383	9660
Breast	569	30	357	212	17070
Heart	270	13	120	150	3510
Ionosphere	351	34	225	126	11934
Liver	345	6	200	145	2070
Pima	768	8	500	268	6144
Sonar	208	60	97	111	12480
Spectf heart	267	44	212	55	11748
Vehicle	846	18	647	199	15228

4　实验结果和讨论

本节将选用UCI数据库[2]的10个公共数据集和2个高炉实际生产数据集为对象来验证本文所提方法的有效性. 除了验证 i-SSLM 模型外, 本文还选取 MMTSSVM 模型[37]、SSLM 模型[34]、经典的 C-SVM 模型[35]和SSVD模型[36]进行对比实验. 所有模型均选用高斯核为核函数, 即

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{\sigma^2}). \tag{12}$$

核参数以及模型中的其它参数采用网格搜索法确定. 在C-SVM和SVDD模型中, 包含2个参数: 惩罚参数C和高斯核参数σ; 在SSLM模型和MMTSSVM模型中, 均包含4个参数: 惩罚参数ν, ν_1, ν_2和核参数σ; 而i-SSLM模型中包含5个参数: 惩罚参数ν, ν_1, ν_2, δ和核参数σ. 为简化寻优, 本文令SSLM, MMTSSVM和i-SSLM模型中的$\nu_1=\nu_2$, 参数C的取值范围为$\{2^{-4}, 2^{-2}, 2^0, 2^2, 2^4\}$[34,37]; 模型SSLM和模型$i$-SSLM中参数$\nu$和$\nu_1$的取值范围为$\{10, 30, 50, 70, 90\}$和$\{0.01, 0.001\}$[34]; 模型MMTSSVM中参数$\nu$和$\nu_1$的取值范围均为$\{0.1, 0.3, 0.5, 0.7, 0.9\}$, 参数$\delta$的取值范围为$\{2^{-5}, 2^{-3}, 2^{-1}\}$, 参数$\sigma$的取值范围为$\{2^{-4}, 2^{-2}, 2^0, 2^2, 2^4\}$[37]. 参数的寻优过程采用五折交叉验证法在训

[2]https://archive-beta.ics.uci.edu/

练集上实施, 寻优结果用测试集进行测试. 为了使结果更具说服力, 本文用5次随机实验的平均结果作为评价标准, 即每次实验随机产生训练集和测试集, 得到最优参数和测试精度, 共进行5次实验, 取5次测试精度的平均值作为最终结果. 所有实验都是在Python软件上进行, 版本为Python 3.8, 硬件配置为Intel(R)Core(TM)i5-8250U CPU@1.60 GHz 1.80 GHz, RAM为8.00 GB.

实验结果用G-means $=\sqrt{\text{Sen}\times\text{Spe}}\times 100\%$进行评价, 其中Sen和Spe分别为灵敏度和特异度, 定义为

$$\text{Sen}(\%)=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100,\ \text{Spe}(\%)=\frac{\text{TN}}{\text{TN}+\text{FP}}\times 100, \tag{13}$$

式中: TP, FP表预测正确和错误的正类样本数; TN, FN表预测正确和错误的负类样本数.

表4给出了某次随机实验训练集上挖掘的各个数据集的圆形知识, 表5报道了这些数据集5次随机实验的平均结果, 其中G-means指标以“平均值±标准差”的形式给出, Time是训练时间的平均值. 作为对比, 本文在表5中同时给出了其他4种模型C-SVM, SVDD, SSLM 和MMTSSVM的对应结果. 从表5可以看出, i-SSLM 在 Arrhythmia, Australian, Liver, Pima, Sonar, Spectf Heart和Vehicle数据集上有更好的分类效果; 而在Heart和Ionoshpere数据集上, MMTSSVM模型效果最好; 在Breast数据集上, SSLM模型有最高的G-means值. 表6统计了每种模型在所有数据集上的G-means综合排名. 显然, i-SSLM具有最高的综合排名. 但从表5的训练时间看, i-SSLM则没有优势, 主要原因是因为知识的引入将会增加模型参数, 从而导致训练时间变长. 这里本文主要关心模型精度的提高, 模型效率的改进将在以后的研究中进行关注.

表 4 10组公共数据集的圆形非线性知识

Table 4 Circular nonlinear knowledge of 10 sets of common datasets

数据集	非线性知识
Arrhythmia	$\|(x^{(18)},x^{(22)})-(-1000,0.5862)\|^2-10^6\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Australian	$\|(x^{(8)},x^{(12)})-(0.5862,1.0939)\|^2-0.5239\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Breast	$\|(x^{(22)},x^{(23)})-(-0.7919,-3.2213)\|^2-14.3012\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Heart	$\|(x^{(6)},x^{(9)})-(-3.2213,7.0658)\|^2-60.3027\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Ionosphere	$\|(x^{(1)},x^{(2)})-(0.5862,-1000)\|^2-10^6\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Liver	$\|(x^{(2)},x^{(3)})-(-1.4586,-3.2213)\|^2-13.9616\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Pima	$\|(x^{(2)},x^{(5)})-(0.2344,0.2344)\|^2-0.1098\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Sonar	$\|(x^{(6)},x^{(28)})-(-0.7919,0.2344)\|^2-1.5813\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Spectf heart	$\|(x^{(32)},x^{(42)})-(0.2344,0.2344)\|^2-0.5956\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$
Vehicle	$\|(x^{(6)},x^{(8)})-(-3.2213,-0.3876)\|^2-11.9435\leqslant 0\Rightarrow \boldsymbol{x}\in\mathcal{X}^+$

为了更好地从统计意义上对比这5个模型, 本文引入Friedman检验对实验结果进行分析. 在所有模型没有显著性区别的零假设下, 计算可得 $\chi_F^2=33.68$ 和 $F_F=47.96$, 其中F_F是服从$(4,36)$自由度的F分布. 这一结果远高于$F(4,36)$在显著性水平$\alpha=0.05$下的临界值2.63和在显著性水平$\alpha=0.01$下的临界值3.89, 表明这 5 种模型具有显著性的差别. 同时, 根据表6可知, i-SSLM平均排名最小, 因此, 可以说明相较于另外4个算法, i-SSLM有更好的实验性能.

4.1 高炉数据集实验

高炉炼铁是钢铁生产工艺的上游工序, 消耗整个工序近70%的输入能量, 对其进行建模和控制一直都是钢铁生产过程中的重要课题. 高炉炉温是评定生铁质量的关键指标, 在实际中因其困难的测量问题, 常常用高炉铁水含硅量作为炉温的指示剂[39]: 铁水含硅量的波动反映了炉温的波动, 其持续上升和下降反映炉缸是向热还是向凉变化. 因此, 对高炉铁水含硅量的正确预测和控制构成了高炉操作指导的基础. 针对这一重要问题, 本文将用i-SSLM模型对高炉铁水含硅量进行分类预测.

本文选用与文献[40]中相同的高炉数据集, 即从国内两座体积约为2500 m^{3}和750 m^{3}的高炉采集的数据, 分别标识为高炉(a)和高炉(b). 前者包含794个样本点, 后者含有800个样本点, 所涉及到的高炉变量个数分别为16个和7个, 具体详见文献[40]. 因高炉是一个强惯性系统, 本文同时考虑这些变量的延迟项对铁水含硅量的影响. 文献[40]利用基于模糊熵的反向选择方法进行特征选择, 即一个特征变量的模糊熵越大, 表明这个变量越不重要. 在特征选择过程中, 高炉训练集首先被分成两部分: 训练集和验证集; 然后, 将所有特征变量输入SVM模型, 通过训练、验证得到高炉(a)和高炉(b)的模型精度(即正确分类的数量与验证集大小的比率)分别为79%和69%; 紧接着, 按照模糊熵

大小顺序依次挑选特征变量, 如果删除一个变量可以使模型精度提高, 则删除该变量, 否则保留; 最后, 挑选出高炉(a)的特征变量, 共42个, 高炉(b)的特征变量共9个. 由于这里选用的高炉、实验数据和文献[40]完全相同, 本文直接选用他们的特征选择结果作为模型输入, 输出的设置也沿用这篇文献的结果, 即对高炉(a)/(b), 硅含量在$[0.41, 1.13]$/$[0.37, 2.20]$视为正类, 标识为$+1$; 硅含量在$[0.13, 0.41)$/$[0.18, 0.37)$视为负类, 标识为-1. 因此, 高炉(a)和高炉(b)可用的正负类样本数量分别为为679, 115和584, 216.

表 5 10组公共数据集实验结果

Table 5 Experimental results from 10 sets of common datasets

数据集	C-SVM G-means/% Time/s	SVDD G-means/% Time/s	SSLM G-means/% Time/s	MMTSSVM G-means/% Time/s	i-SSLM G-means/% Time/s
Arrhythmia	31.77±9.99 0.31	44.34±2.88 0.16	60.65±1.40 0.23	50.80±5.50 0.13	**63.28±2.26**[1] 2.58
Australian	80.77±4.47 0.57	72.61±2.14 0.53	85.26±1.66 0.83	85.48±0.61 0.43	**86.70 ± 0.98** 5.58
Breast	91.89 ± 3.17 0.64	92.29 ± 3.53 0.74	**95.38 ±0.40** 0.80	92.81±3.43 0.57	93.86 ±2.58 5.82
Heart	66.21 ±4.36 0.08	64.49 ±3.15 0.08	72.68 ±8.14 0.13	**77.55±4.46** 0.06	76.34 ±4.96 0.81
Ionosphere	80.31 ±3.84 0.32	88.89±4.41 0.25	90.34±3.18 0.29	**94.16±2.81** 0.26	92.25±1.37 2.80
Liver	21.04 ±21.16 0.31	54.67 ± 6.45 0.22	59.00 ± 8.04 0.33	59.32±7.06 0.18	**63.76 ± 2.46** 0.91
Pima	45.59 ±10.50 1.80	60.15 ±4.19 1.26	70.76 ±3.30 1.66	71.34±2.40 1.17	**73.47 ±2.00** 6.42
Sonar	45.96 ±26.27 0.09	58.42 ±3.47 0.05	66.03 ±8.33 0.09	64.07±6.74 0.04	**69.15 ±5.97** 0.55
Spectf heart	28.41 ±16.76 0.44	53.63 ±4.59 0.29	72.57 ±3.49 0.38	70.15±2.71 0.23	**74.28±2.77** 2.16
Vehicle	91.73±1.69 2.26	87.13±4.45 2.02	95.61±1.58 1.70	93.51±2.22 1.88	**96.44 ± 0.74** 25.87

注: 黑体表示每组最优精度.

表 6 5种模型在公共数据集上的综合排名

Table 6 Comprehensive ranking of 5 models on public datasets

数据集	C-SVM	SVDD	SSLM	MMT-SSVM	i-SSLM
Arrhythmia	5	4	2	3	1
Australian	4	5	3	2	1
Breast	5	4	1	3	2
Heart	5	4	3	1	2
Ionosphere	5	4	3	1	2
Liver	5	4	3	2	1
Pima	5	4	3	2	1
Sonar	5	4	2	3	1
Spectf heart	5	4	2	3	1
Vehicle	4	5	2	3	1
Average rank	4.8	4.2	2.4	2.3	1.3

高炉实验与公共数据集实验设置有所不同, 训练集和测试集通过下述 5 种方式构造: 从正类样本种随机选取 70% 样本, 负类样本中分别随机选择 10%～50%的样本构成训练集; 剩下的样本作为测试集. 对于每种模式构造的样本集, 均进行5次实验, 最终结果取5次测试结果的平均值; 参数训练时仍采用5层交叉验证. 为了更好地展现i-SSLM模型在高炉实验中的优势, 本文同时对比了MMTSSVM模型. 它们的参数寻优范围, 除了i-SSLM模型的参数σ寻优范围为$\{2^0, 2^2, 2^4\}$外, 其他和第3.3节公共数据集相同.

表 7 展示了利用算法1在高炉(a)和高炉(b)中针对不同比例的负类样本点挖掘的非线性知识. 以50%比例的负类样本点为例, 图1展示了挖掘的圆形知识对样本的分类情况. 从图中可以看出, 所有的负类样本均在圆外, 这意味着圆内所有样本点均为正类, 包括数据集原有的部分正类样本点$\mathcal{I}_1$(三角形)及知识离散化产生的部分正类样本点$\mathcal{I}_2$(星形). 这些样本在i-SSLM模型训练之前就已经明确需要分为正类, 因此模型具有一定的可解释性. 同时因为$\mathcal{I}_2$样本增加了新的数据信息, 模型的精度也会随之提高. 进一步观察表7所列知识(以高炉(a)为例), 所涉及变量为x^{14}(喷煤)和z(硅含量), 这基本与高炉工艺原理较吻合. 因为在高

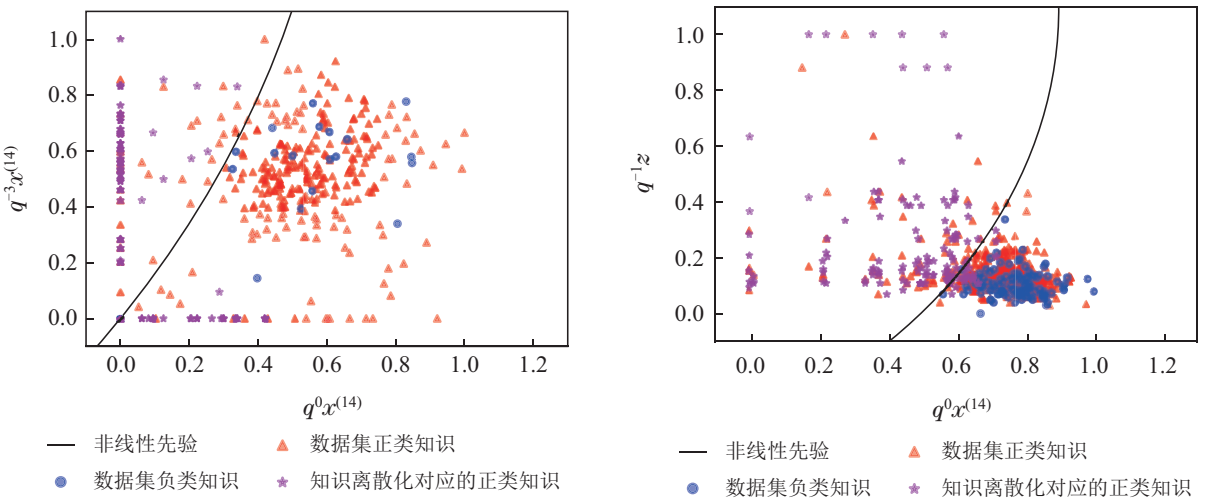
炉冶炼过程中, 对硅含量的影响因素有很多, 但其历史值(高炉强惯性特征)和喷煤量是两个非常重要的因素; 同时, 高炉是个滞后系统, 喷煤量的调整作用会在一定时间之后显现出来. 所以, 本文所挖掘的知识和高炉实际知识是相对应的, 这也反向说明了本文知识挖掘算法的有效性.

表 7 高炉数据挖掘的非线性知识

Table 7 Nonlinear knowledge of blast furnace data mining

BF	Proportion%	非线性知识
(a)	10	$\|(q^{-3}x^{(14)}, q^{-1}z) - (1.0939, 0.5862)\|^2 - 0.7794 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	20	$\|(q^{0}x^{(14)}, q^{-4}x^{(14)}) - (-0.7919, 1.0939)\|^2 - 1.8237 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	30	$\|(q^{0}x^{(14)}, q^{-4}x^{(14)}) - (-0.7919, 1.0939)\|^2 - 1.8237 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	40	$\|(q^{0}x^{(14)}, q^{-3}x^{(14)}) - (-3.2213, 2.1209)\|^2 - 14.875 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	50	$\|(q^{0}x^{(14)}, q^{-3}x^{(14)}) - (-3.2213, 2.1209)\|^2 - 14.875 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
(b)	10	$\|(q^{0}x^{(8)}, q^{-1}z) - (0.5862, 0.2344)\|^2 - 0.0283 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	20	$\|(q^{0}x^{(14)}, q^{-1}x^{(14)}) - (0.5862, 0.5862)\|^2 - 0.0281 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	30	$\|(q^{0}x^{(14)}, q^{-1}z) - (-0.7919, 1.0939)\|^2 - 2.9794 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	40	$\|(q^{0}x^{(14)}, q^{-1}z) - (-0.7919, 1.0939)\|^2 - 2.8600 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$
	50	$\|(q^{0}x^{(14)}, q^{-1}z) - (-0.7919, 1.0939)\|^2 - 2.8600 \leqslant 0 \Rightarrow \boldsymbol{x} \in \mathcal{X}^+$

注: BF 表示 Blast furnace; Proportion 表示训练集中采用的负类点比例.



(a) 高炉(a)

(b) 高炉(b)

图 1 高炉数据负类样本50%比例下挖掘知识的分类边界

Fig. 1 Mining the classification boundary of knowledge under the negative sample of blast furnace data is 50% ratio

表8给出了 i-SSLM 模型和 MMTSSVM 模型在高炉数据集上的实验结果, 包括G-means, Sen和Spe这3种指标的5次平均结果, 同样以“平均值±标准差”的形式给出. 由表8可知, 对于高炉(a)数据集, i-SSLM比MMTSSVM有更高的G-means和特异度; 当负类点比例为30%和40%时, i-SSLM的灵敏度略低于MMTSSVM. 对于高炉(b)数据集, 虽然i-SSLM的灵敏度和特异度在部分负类点比例上比MMTSSVM低, 但是总体上, i-SSLM比MMTSSVM有更高的G-means. 综合而言, 在高炉数据集上i-SSLM比MMTSSVM展现了更好的性能.

5 结论与展望

本文针对SSLM模型具有处理非平衡数据能力但缺乏可解释性等问题, 提出从SSLM黑箱模型的输入端集成先验知识, 以增强黑箱模型的可解释性和精度. 论文取得的结果如下: 1) 提出了一种从数据中挖掘2维圆形非线性知识的方法, 圆内可最大限度地包含正类样本, 而圆外则包含所有负类样本; 2) 开发了一种离散化圆形区域知识的算法, 离散化后的数据点, 不仅包含产生知识的原有数据样本点, 还产生一些新的数据样本点; 3) 将挖掘的圆形先验知识以不等式约束的形式集成至SSLM模型, 构造了i-SSLM模型, 后者因确保圆内样本点为正类样本因而具有可解释性, 又因离散化圆形知识增加了新的样本点因而同时具有更高的精度; 4) 10个公共样本集和2个实际高炉数据集例证了i-SSLM 模型在可解释性和精度上优于传统的C-SVM, SVDD, SSLM, MMTSSVM4种模型.

表 8 高炉数据的实验结果

Table 8 Experimental results of blast furnace data

BF	Algorithm	Indicator%	10%	20%	30%	40%	50%
(a)	i-SSLM	G-means	65.25 ± 2.48	68.98 ± 3.45	71.43 ± 2.42	69.50 ± 1.43	65.79 ± 4.09
		Sen	73.43± 8.90	73.63 ± 9.17	74.71 ± 6.99	72.06 ± 4.06	70.78 ± 2.21
		Spe	58.45± 5.25	65.65 ± 11.10	68.64 ± 5.28	67.25 ± 5.19	61.40 ± 8.13
	MMTSSVM	G-means	54.87±9.19	62.69 ±2.84	57.58 ±6.24	65.35 ± 1.97	63.02 ±3.74
		Sen	68.53± 18.03	67.55 ±8.44	76.27 ±11.34	77.06 ± 4.14	68.92 ± 9.93
		Spe	49.32± 21.95	58.91± 7.85	45.68 ±15.81	55.65± 5.29	58.95±11.54
(b)	i-SSLM	G-means	60.90 ± 2.16	62.30 ± 2.22	61.84 ± 4.77	64.04 ± 2.90	64.42 ± 3.75
		Sen	67.09 ± 6.72	67.09 ± 10.67	66.40 ± 8.10	65.60 ± 5.59	64.91± 8.85
		Spe	55.67 ± 5.71	59.42 ± 12.30	58.68 ± 11.61	62.92 ± 7.18	64.44 ± 5.62
	MMTSSVM	G-means	57.65± 3.27	61.48 ± 2.50	60.19 ± 4.66	61.78± 3.58	62.42 ± 3.03
		Sen	60.23 ±10.46	58.63 ± 7.91	67.09 ± 4.21	65.60 ± 3.84	66.40 ± 5.31
		Spe	56.29 ± 9.09	65.32 ± 8.43	54.30 ± 7.78	58.46 ± 6.90	59.07 ± 7.15

尽管i-SSLM模型展现了良好的性能, 但仍有较大改进空间: a) 挖掘的圆形知识只涉及两个特征变量, 未来的研究可考虑涉及更多特征变量、更高维的球形或球体先验知识, 以挖掘更准确的先验知识; b) 挖掘的知识仅为判定样本为正类的先验知识, 对于二分类问题, 可进一步挖掘负类先验知识集成至黑箱模型, 全面增强黑箱模型的可解释性; c) 知识的引入使得i-SSLM模型包含更多的参数, 导致后续的参数寻优和模型训练均耗费更长时间, 如何提高i-SSLM模型的学习效率是未来研究的重点.

参考文献:

[1] JAEWON J, JEONGHUN K, SANGWON P, et al. Machine learning-based automatic estimation of cortical atrophy using brain computed tomography images. *Scientific Reports*, 2022, 12: 14740.

[2] AVIAD L, PRATUL P S, ANDREW A C, et al. Gravitationally lensed black hole emission tomography. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA: IEEE, 2022: 19841 – 19850.

[3] ZHANG Z Y, LU Y F, WANG X J, et al. A data-based compact high-order volterra model for complex blast furnace system. *IEEE Transactions on Industrial Informatics*, 2022, 18(9): 5827 – 5837.

[4] LAURA V R, SEBASTIAN M, KATHARINA B, et al. Informed machine learning–A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 614 – 633.

[5] ZHU Hufei, DING Zihao, YANG Yongliang, et al. Two-stage sparse representation objective tracking algorithm in reproducing kernel Hilbert space. *Control Theory & Applications*, 2022, 39(4): 730 – 740.
(朱虎飞, 丁子豪, 杨永亮, 等. 再生核Hilbert空间中的两阶段稀疏表示目标跟踪算法. 控制理论与应用, 2022, 39(4): 730 – 740.)

[6] ZHANG Qing, YAN Xuefeng. Support vector regression algorithm that combines probability distributions and monotonicity. *Control Theory & Applications*, 2017, 34(5): 671 – 676.
(张青, 颜学峰. 融合概率分布和单调性的支持向量回归算法. 控制理论与应用, 2017, 34(5): 671 – 676.)

[7] TANG Jian, CHAI Tianyou, CONG Qiumei, et al. Modeling of mill load parameters with selective fusion of multi-scale cylinder vibration spectrum. *Control Theory & Applications*, 2015, 32(12): 1582 – 1591.
(汤健, 柴天佑, 丛秋梅, 等. 选择性融合多尺度筒体振动频谱的磨机负荷参数建模. 控制理论与应用, 2015, 32(12): 1582 – 1591.)

[8] HOU B J, ZHOU Z H. Learning with interpretable structure from gated rnn. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 31(7): 2267 – 2279.

[9] DAVID M, BART B, TONY V G. Decompositional rule extraction from support vector machines by active learning. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(2): 178 – 191.

[10] JAN C, JACEK M Z. Extracting rules from neural networks as decision diagrams. *IEEE Transactions on Neural Networks*, 2011, 22(12): 2435 – 2446.

[11] THUAN Q, HUYNH, JAMES A R. Guiding hidden layer representations for improved rule extraction from neural networks. *IEEE Transactions on Neural Networks*, 2010, 22(2): 264 – 275.

[12] NAHLA B, ANDREW P B. Rule extraction from support vector machines: A review. *Neurocomputing*, 2010, 74(1/ 3): 178 – 190.

[13] ZHOU Z J, CAO Y, HU G Y, et al. New health-state assessment model based on belief rule base with interpretability. *Science China Information Sciences*, 2021, 64(172214): 1 – 15.

[14] LAUER F, BLOCH G. Incorporating prior knowledge in support vector regression. *Machine Learning*, 2008, 70(1): 89 – 118.

[15] QU Y J, HU B G. Generalized constraint neural network regression model subject to linear priors. *IEEE Transactions on Neural Networks*, 2011, 22(12): 2447 – 2459.

[16] RAFAEL V B, ARTUR D G, LUIS C L. Learning and representing temporal knowledge in recurrent networks. *IEEE Transactions on Neural Networks*, 2011, 22(12): 2409 – 2421.

[17] LAURA R, CHANDAN S, WILLIAM M, et al. Interpretations are useful: Penalizing explanations to align neural networks with prior knowledge. *ArXiv Preprint*, 2020: arXiv:1909.13584.

[18] PARTHA N, FEDERICO G, TOMASO P. Incorporating prior information in machine learning by creating virtual examples. *Proceedings of the IEEE*, 1998, 86(11): 2196 – 2209.

[19] OLIVIER C, BERNHARD S. Incorporating invariances in non-linear support vector machines. *Advances in Neural Information Processing Systems*. Canada: NIPS, 2001: 609 – 616.

[20] LAUER F, BLOCH G. Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing*, 2008, 71(7/9): 1578 – 1594.

[21] DENNIS D, BERNHARD S. Training invariant support vector machines. *Machine Learning*, 2002, 46(1): 161 – 190.

[22] HAASDONK B, KEYSERS D. Tangent distance kernels for support vector machines. *The 16th International Conference on Pattern Recognition (ICPR)*. Quebec, Canada: IEEE, 2002, 2: 864 – 868.

[23] UTKIN L V. An imprecise extension of SVM-based machine learning models. *Neurocomputing*, 2019, 331: 18 – 32.

[24] ALEXEI P, SAMY B. Invariances in kernel methods: From samples to objects. *Pattern Recognition Letters*, 2006, 27(10): 1087 – 1097.

[25] HAASDONK B, VOSSEN A, BURKHARDT H. Invariance in kernel methods by haar-integration kernels. *The 14th Scandinavian Conference on Image Analysis*. Joensuu, Finland: Springer, 2005, 3540: 841 – 851.

[26] RISI K, TONY J. A kernel between sets of vectors. *In Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. Washington, DC, USA: PMLR, 2003, 361 – 368.

[27] LIOR W, AMNON S. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 2003, 4(10): 913 – 931.

[28] WANG L, GAO Y, KAP L C, et al. Retrieval with knowledge-driven kernel design: An approach to improving svm-based cbir with relevance feedback. *In the 10th IEEE International Conference on Computer Vision (ICCV'05)*. Beijing, China: IEEE, 2005, 2: 1355 – 1362.

[29] GLENN F, OLVI M, JUDE S. Knowledge-based support vector machine classifiers. *Advances in Neural Information Processing Systems*. Canada: NIPS, 2002: 537 – 544.

[30] ZHAO J, XU Y, FUJITA H. An improved non-parallel universum support vector machine and its safe sample screening rule. *Knowledge-Based Systems*, 2019, 170: 79 – 88.

[31] PU G, WANG L, SHEN J, et al. A hybrid unsupervised clustering-based anomaly detection method. *Tsinghua Science and Technology*, 2020, 26(2): 146 – 153.

[32] CHEN S H, GAO C H. Linear priors mined and integrated for transparency of blast furnace black-box svm model. *IEEE Transactions on Industrial Informatics*, 2020, 16(6): 3862 – 3870.

[33] CHEN S H, GAO C H, ZHANG P. Incorporation of data-mined knowledge into black-box svm for interpretability. *ACM Transactions on Intelligent Systems and Technology*, 2022, 14(1): 1 – 22.

[34] WU M R, YE J P. A small sphere and large margin approach for novelty detection using training data with outliers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(11): 2088 – 2092.

[35] VAPNIK V. *The Nature of Statistical Learning Theory*. New York, USA: Springer, 1995.

[36] DAVID M J T, ROBERT P W D. Support vector data description. *Machine Learning*, 2004, 54(1): 45 – 66.

[37] XU Y T. Maximum margin of twin spheres support vector machine for imbalanced data classification. *IEEE Transactions on Cybernetics*, 2016, 47(6): 1540 – 1550.

[38] OLVI L M, EDWARD W W. Nonlinear knowledge-based classification. *IEEE Transactions on Neural Networks*, 2008, 19(10): 1826 – 1832.

[39] ZHOU P, ZHANG R, XIE J, et al. Data-driven monitoring and diagnosing of abnormal furnace conditions in blast furnace ironmaking: An integrated PCA-ICA method. *IEEE Transactions on Industrial Electronics*, 2021, 68(1): 622 – 631.

[40] GAO C H, GE Q H, JIAN L. Rule extraction from fuzzy-based blast furnace svm multiclassifier for decisionmaking. *IEEE Transactions on Fuzzy Systems*, 2013, 22(3): 586 – 596.

作者简介:

陆思洁 硕士研究生, 目前研究方向为 机器学习, E-mail: 21935084@zju.edu.cn;

范 崚 博士研究生, 目前研究方向为 机器学习, E-mail: fandi@zju.edu.cn;

渐 令 教授, 博士, 博士生导师, 目前研究方向为 机器学习、最优化理论与应用, E-mail: bebetter@upc.edu.cn;

郜传厚 教授, 博士, 博士生导师, 目前研究方向为 数学系统生物学、热力学系统控制、机器学习和优化, E-mail: gaochou@zju.edu.cn.