

DOI: 10.11992/tis.201805010

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190115.1151.002.html>

## 一种双优选的半监督回归算法

程康明<sup>1</sup>, 熊伟丽<sup>1,2</sup>

(1. 江南大学 物联网工程学院, 江苏 无锡 214122; 2. 江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

**摘 要:** 针对一些工业过程中存在的有标签样本少, 而传统的半监督学习无法保证对无标签样本准确预测的问题, 提出一种双优选的半监督回归算法。首先, 确定有标签样本密集区中心, 并计算无标签样本与该中心的相似度, 实现对无标签样本的优选, 同时根据有标签样本间相似度优选有标签样本; 然后, 利用高斯过程回归方法对选出的有标签样本建立辅学习器, 以对优选出的无标签样本预测标签; 最后, 利用这些伪标签样本提升主学习器的预测效果。通过数值例子以及实际脱丁烷塔过程数据进行建模仿真, 证明了所提方法在有标签样本较少的情况下有良好的预测性能。

**关键词:** 无标签样本; 优选; 半监督回归; 样本密集区中心; 相似度; 高斯过程回归; 辅学习器; 主学习器; 脱丁烷塔过程; 预测性能

**中图分类号:** TP274    **文献标志码:** A    **文章编号:** 1673-4785(2019)04-0689-08

中文引用格式: 程康明, 熊伟丽. 一种双优选的半监督回归算法 [J]. 智能系统学报, 2019, 14(4): 689-696.

英文引用格式: CHENG Kangming, XIONG Weili. A dual-optimal semi-supervised regression algorithm[J]. CAAI transactions on intelligent systems, 2019, 14(4): 689-696.

## A dual-optimal semi-supervised regression algorithm

CHENG Kangming<sup>1</sup>, XIONG Weili<sup>1,2</sup>

(1. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; 2. Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China)

**Abstract:** Aiming at the problem that there are few label samples in some industrial processes and that the traditional semi-supervised learning cannot guarantee the accurate prediction of unlabeled samples, a dual-optimal semi-supervised regression algorithm is proposed in this paper. First, in this method, the center of the label-concentrated area is found, and the similarity between unlabeled samples and the center is calculated, and therefore, the unlabeled samples are optimized. At the same time, the labeled samples are selected according to similarity between the unlabeled samples and the center of the dense area. Second, by employing the Gaussian process regression method, an auxiliary learner is established according to the selected labeled sample, and then the labels of the selected unlabeled samples are predicted by the auxiliary learner. Finally, the performance of the main learner is improved with these pseudo-label samples. Through a simulation of the numerical case and the actual debutanizer process, the proposed method is verified to have a good prediction performance when the labeled samples are few.

**Keywords:** unlabeled samples; select; semi-supervised regression; center of sample dense area; similarity; Gaussian process regression; auxiliary learner; main learner; debutanizer process; prediction performance

针对一些工业过程存在有标签样本数据少且获取代价大, 而无标签样本数据大量存在的情

况, 研究利用少量有标签样本和大量的无标签样本来提高学习性能的半监督学习受到了密切关注<sup>[1-3]</sup>。如今, 半监督思想已经应用于各行各业, 根据半监督学习目的将其大致分为 3 类<sup>[4]</sup>: 半监督聚类、半监督分类和半监督回归。其中, 半监

收稿日期: 2018-05-09. 网络出版日期: 2019-01-16.

基金项目: 国家自然科学基金项目 (61773182, 60712228); 江苏省自然科学基金项目 (BK20170198).

通信作者: 熊伟丽. E-mail: [greenpre@163.com](mailto:greenpre@163.com).

督聚类与半监督分类的研究很多,而半监督回归的研究相对较少。

半监督回归根据学习方法的不同大致分为两类<sup>[5-6]</sup>:第一类为利用流行学习的半监督回归算法;第二类为协同训练算法。其中,流行学习方面的研究有杨剑等<sup>[7]</sup>提出的一类广义损失函数的 Laplacian 半监督回归方法,该方法充分利用了有标签样本的结构信息来提高回归估计的精度。协同训练的典型代表为 Zhou 等<sup>[8]</sup>在 2005 年提出的基于协同训练的半监督回归算法,该方法通过建立两个学习器,以交互学习的方式来利用无标签样本,达到提高回归精度的目的,但是其对有标签样本两个冗余视图的假设,使得该方法在某些场合的应用受到限制。另外,一些其他的半监督回归方法也取得了不错的效果。如程玉虎等<sup>[9]</sup>提出的一种 Help-Training 的半监督支持向量回归算法,通过置信度评估选出了信任的无标签样本。盛高斌等<sup>[10]</sup>提出的基于半监督回归的选择性集成算法,考虑了如何通过有标签样本的合理利用提高无标签样本预测的准确性。

上述方法有的从有标签样本角度出发,有的从无标签样本角度出发,本文为了更准确地利用无标签样本,同时考虑对无标签样本和有标签样本的筛选,通过定义两个优选准则,提出一种双优选的半监督回归算法。该方法首先筛选无标签样本,降低引入预测误差的可能性,同时筛选有标签样本,获得一个更有针对性的有标签样本集;然后利用高斯过程回归(GPR)方法对选出的有标签样本建立辅学习器,以对选出的无标签样本预测标签;最后利用这些伪标签样本和初始有标签样本集通过 GPR 方法建立主学习器,从而提升主学习器预测效果。通过两个仿真实验,验证了本文所提方法的有效性。

## 1 高斯过程回归

GPR 是一种基于统计学习理论的非参数概率模型,适合处理高维度、小样本及非线性等数据的建模问题<sup>[11]</sup>。

高斯过程回归算法流程简述如下,给定训练样本集  $\{X, y\}$ , 其中  $X = \{x_i \in \mathbb{R}^D\}$ ,  $y = \{y_i \in \mathbb{R}\}$ ,  $i = 1, 2, \dots, n$  分别代表  $D$  维的输入和输出数据,通常观测值  $y_i$  与输入  $x_i$  应满足  $y_i = f(x_i) + \varepsilon$ , 其中  $f$  为未知函数形式,高斯噪声  $\varepsilon \sim N(0, \sigma^2)$ 。对于一个新的输入  $x^*$ , 相应的预测值  $y^*$  也应满足高斯分布,其均值和方差如式(1)、式(2)所示:

$$\mu(y^*) = k^T(x^*)K^{-1}y \quad (1)$$

$$\sigma^2(y^*) = C(x^*, x^*) - k^T(x^*)K^{-1}k(x^*) \quad (2)$$

式中:  $k(x^*)$  为训练样本与测试样本的协方差矩阵;  $K$  为训练样本的自协方差矩阵;  $C$  为测试样本自协方差。

协方差函数是 GPR 方法中的重要组成部分。GPR 方法可以选择多种协方差函数,只需保证对任意输入能够得到一个非负定协方差矩阵。本文选择常用的高斯协方差函数,定义为

$$k(x_i, x_j) = \nu \exp \left[ -\frac{1}{2} \sum_{d=1}^D \omega_d (x_i^d - x_j^d)^2 \right] \quad (3)$$

式中:  $\nu$  为先验知识的总体度量;  $\omega_d$  表示每个成分  $x^d$  的重要性程度。

利用式(1)~(3)确定均值和方差后,对测试样本输入  $x^*$ , 即可预测其输出  $y^*$ 。详细的 GPR 建模过程可参考文献[12]。

## 2 基于双优选的半监督回归策略

本文提出的双优选策略,通过定义两个优选准则来综合考虑无标签样本的筛选,有标签样本筛选以建立辅学习器两方面问题,进而达到准确预测无标签样本的目的。

### 2.1 双优选准则

优选准则 1: 给定一个阈值  $\theta_1$ , 利用马氏距离<sup>[13]</sup>来度量无标签样本  $x_i'$  与有标签样本密集区中心  $C$  的相似度,若  $x_i'$  与  $C$  的距离小于  $\theta_1$ , 则  $x_i'$  满足优选条件。

优选准则 2: 给定一个阈值  $\theta_2$ , 利用马氏距离来度量有标签样本间的相似度,统计有标签样本  $x_i$  与其余有标签样本的马氏距离小于  $\theta_2$  的有标签样本数量  $m$ , 若  $m$  不小于 2, 则  $x_i$  满足优选条件。

上述两个优选准则,优选准则 1 通过马氏距离选出以有标签样本密集区中心为球心  $\theta_1$  为半径的球域内所有无标签样本点,利用该准则有利于选出能够被准确预测的无标签样本,剔除了可能带来较大预测噪声的无标签样本。优选准则 2 采用了 Knoor 等<sup>[14]</sup>提出的基于距离的离群点的定义,通过马氏距离度量的相似度,剔除了阈值  $\theta_2$  限定下的离群点,从而提升辅学习器的预测精度。有标签样本  $x_i$  与  $x_j$  之间马氏距离  $d(x_i, x_j)$  计算公式如下:

$$d(x_i, x_j) = \text{sqrt}[(x_i - x_j)^T S^{-1} (x_i - x_j)] \quad (4)$$

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (5)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

式中:  $S$  为有标签样本协方差矩阵;  $n$  为变量个数;  $\bar{x}$  为有标签样本均值。

另外, 优选准则 1 中  $x'_i$  与  $C$  的距离  $d_i$  同样可用式 (4)~(6) 获得。

## 2.2 无标签样本筛选

为了选出能够被准确预测的无标签样本, 利用优选准则 1 对无标签样本进行筛选, 选出满足条件的无标签样本。而该准则能否做出有效筛选, 取决于样本密集区中心是否准确, 尤其在有标签样本数目比较少时, 此时样本密集区中心对离群点更敏感。针对这种情况, 提出一种新的寻找有标签样本密集区中心的方法。该方法首先找出有标签样本中属于样本密集区的样本, 然后对这部分属于样本密集区有标签样本求各属性均值, 得到样本密集区中心  $C$ 。其具体操作如下, 依次计算每一个有标签样本与周围有标签样本间的马氏距离  $d(x_i, x_j)$ , 然后判断  $d(x_i, x_j)$  是否小于阈值  $\theta_3$ , 统计  $x_i$  周围满足条件的有标签样本数量  $h$ , 若  $h$  大于 1, 则  $x_i$  为有标签样本密集区样本, 最后用所得样本密集区样本确定样本密集区中心。在确保准则 1 的筛选效果后, 再利用其筛选无标签样本。具体伪代码如算法 1 所示。

### 算法 1 无标签样本筛选

**输入** 有标签样本集  $S$  (由自变量  $x_i$  与因变量  $y_i$  组成), 无标签样本集  $M$  (只包含自变量  $x'_i$ ), 阈值  $\theta_1$  和  $\theta_3$ , 计数变量  $h$

**输出** 选出的无标签样本集  $M_1$

```

1)  $h \leftarrow 0$ 
2) for  $i = 1$  to 50 do
3) for  $j = i + 1$  to 50 do
4) 利用公式 (1)~(3) 计算两样本间的马氏距离  $d(x_i, x_j)$ 
5) if  $(d(x_i, x_j) < \theta_1)$  then
6)  $h \leftarrow h + 1$ 
7) End if
8) End for
9) if  $(h > 1)$  then
10)  $A \leftarrow x_i$ 
14) End if
15)  $h \leftarrow 0$ 
13) End for
14)  $C(k) \leftarrow \frac{1}{l} \sum_{m=1}^l A(m, k)$   $l$  为  $A$  中样本数量,  $k$  为样本维度
15) for  $i = 1$  to 500 do

```

16) 利用公式 (1)~(3) 计算样本与密集区中心  $C$  的马氏距离  $d_i$

17) if  $(d_i < \theta_3)$  then

18)  $M_1 \leftarrow x'_i$

19) End if

20) End for

## 2.3 辅学习器建立

为了建立一个更有针对性的辅学习器, 已知无标签样本集  $M_1$  分布于有标签样本密集区中心  $C$  周围的情况下, 利用优选准则 2 筛选出与  $M_1$  特性更相似的有标签样本, 然后根据选出的有标签样本建立辅学习器, 实现对  $M_1$  更准确地预测。具体伪代码如算法 2 所示。

### 算法 2 辅学习器建立

**输入** 有标签样本集  $S$  (由自变量  $x_i$  与因变量  $y_i$  组成), 阈值  $\theta_2$ , 计数变量  $h$

**输出** 辅学习器  $f_1$

```

1)  $h \leftarrow 0$ 
2) for  $i = 1$  to 50 do
3) for  $j = i + 1$  to 50 do
4) 利用公式 (1)~(3) 计算两样本间的马氏距离  $d(x_i, x_j)$ 
5) if  $(d(x_i, x_j) < \theta_2)$  then
6)  $h \leftarrow h + 1$ 
7) End if
8) End for
9) if  $(h > 1)$  then
10)  $A_1 \leftarrow x_i$ 
11)  $B_1 \leftarrow y_i$ 
12) End if
13)  $h \leftarrow 0$ 
14) End for
15)  $f_1 \leftarrow \text{GPR}(A_1, B_1)$ 

```

## 2.4 基于双优选的半监督算法流程

综合考虑无标签样本和有标签样本的选择, 提出了一种带双优选过程的建模方法, 建模流程如图 1 所示。具体建模步骤如下:

1) 根据优选准则 1 筛选出无标签样本, 得到无标签样本集  $M_1$ 。

2) 根据优选准则 2 选出有标签样本, 建立一个更有针对性的辅学习器  $f_1$ 。

3) 利用辅学习器  $f_1$  对无标签样本集  $M_1$  预测其标签, 将所得的伪标签样本集  $S_1$  添加到初始有标签样本集  $S$  中, 最后再利用 GPR 方法建立主学习器。

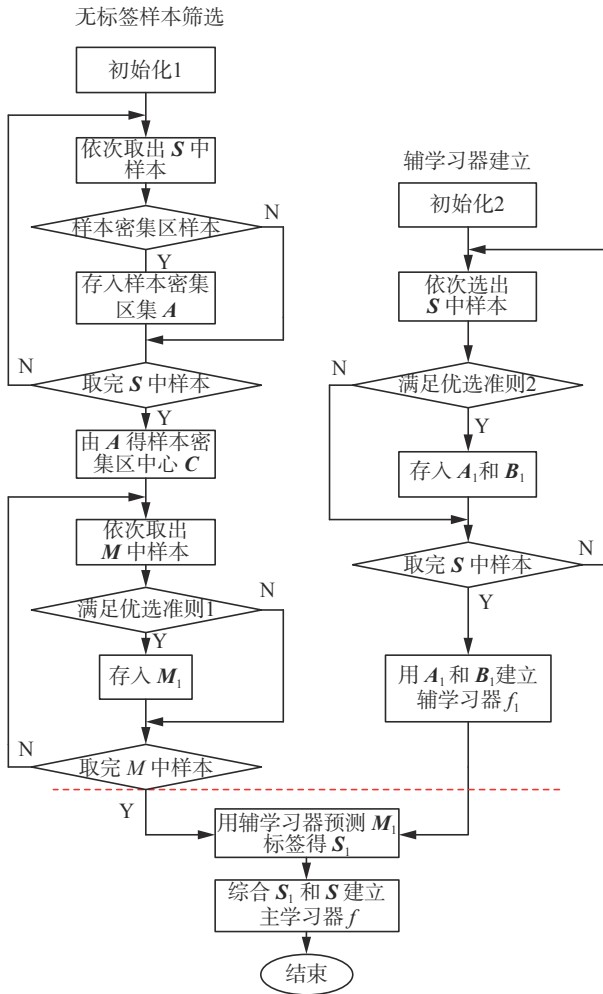


图1 总体算法步骤

Fig. 1 Overall algorithm steps

### 3 数值仿真实验

为了验证所提双优选方法的有效性,采用文

献 [15] 中的非线性函数进行数值仿真:

$$x_1(t+1) = \left( \frac{x_1(t)}{1+x_1^2(t)} + 1 \right) \sin(x_2(t))$$

$$x_2(t+1) = x_2(t) \cos(x_2(t)) + x_1(t) \exp\left(-\frac{x_1^2(t)+x_2^2(t)}{8}\right) + \frac{u^3(t)}{1+u^2(t)+0.5\cos(x_1(t)+x_2(t))}$$

$$y(t) = \frac{x_1(t)}{1+0.5\sin(x_2(t))} + \frac{x_2(t)}{1+0.5\sin(x_1(t))} + \varepsilon(t) \quad (7)$$

式中:  $x_1(t)$  与  $x_2(t)$  为系统状态;  $u(t)$ 、 $y(t)$  和  $\varepsilon(t)$  分别为系统的输入、输出和白噪声。由于系统状态变量是不可测的,只能利用可测的输入输出信息来预测系统输出  $y(t)$ , 仿真模型结构如下:

$$y(t) = f\left(\begin{matrix} y(t-1), y(t-2), y(t-3) \\ u(t-1), u(t-2), u(t-3) \end{matrix}\right) \quad (8)$$

同时为了解决由于训练样本与测试样本过于相似而带来的对建模效果评价的可信度问题,仿真中通过两种不同方式分别产生训练样本和测试样本。对于训练样本,首先产生  $[-2,2]$  的随机信号  $u(t)$  与  $[-2,2]$  的随机信号  $\varepsilon(t)$ ,然后将这两组信号作用于系统,最终得到 600 组数据作为训练样本;对于测试样本,首先令  $u(t) = \sin(0.2\pi t) + \sin(0.08\pi t)$ ,产生  $[-2,2]$  的随机信号  $\varepsilon(t)$ ,然后这两组信号作用于系统,最终得到 200 组数据作为测试样本。600 组训练样本中选取 50 组作为有标签样本,另外 550 组作为无标签样本。

为了更直观地说明双优选方法的必要性,对数值仿真产生的训练样本数据进行了直方图统计,分别统计了变量在每一维度上的分布特点,具体如图 2 所示,其中横坐标为样本分布区间,纵坐标为样本出现的频数。

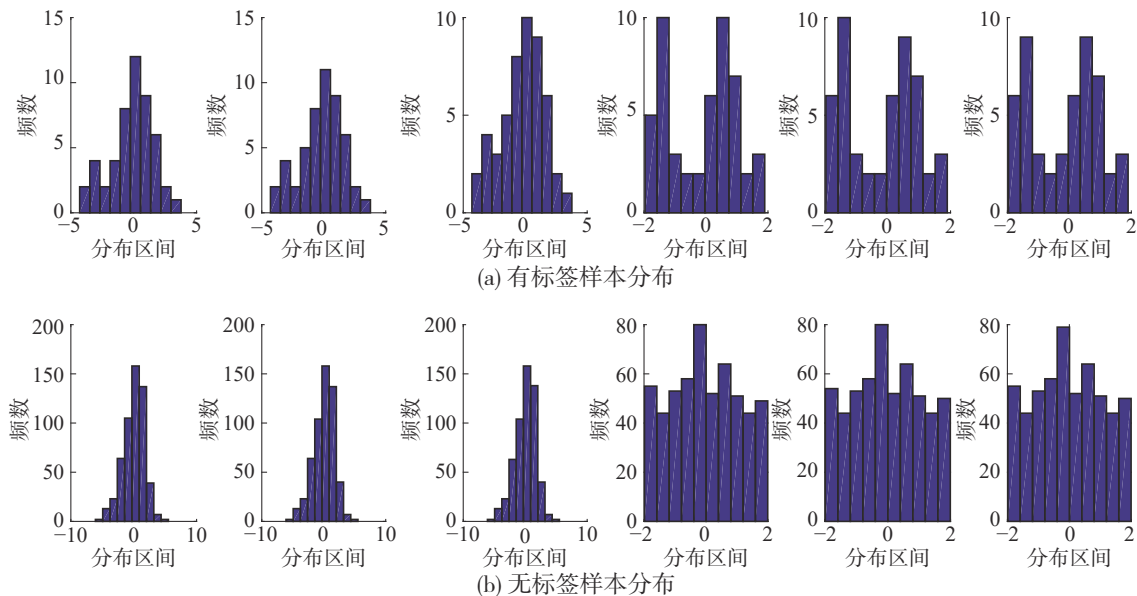


图2 有标签样本与无标签样本的直方图分布

Fig. 2 Histogram distribution about labeled and unlabeled samples



在图2中,图(a)为50组有标签样本辅助变量分布的直方图统计,图(b)为550组无标签样本每一维度分布的直方图统计。因为大量无标签样本更能够反映过程的整体特点,由图2(a)和图2(b)前3幅图的对比,发现无标签样本分布大致成一种正态分布,而图2(a)有标签样本的分布中,−5~0这段趋势明显异于整体的正态分布,可知部分初始有标签样本是游离在整体特性之外的。为保证准确性,需使建模样本与预测样本分布特性尽量一致。由图2(a)后3幅图发现,有标签样本分布在−1~1这段的数量相对较少,约占全部有标签样本的1/2,而由图2(b)后3幅图发现,无标签样本分布在−1~1这段的数量明显更多,约占全部无标签样本的3/4。由上述对比可知,仅靠现有的有标签样本,难以准确预测所有的无标签样本。因此,想要利用无标签提升建模效果,需要根据优选准则1筛选出能够被准确预测的无标签样本,剔除那些可能会带来大量噪声的无标签样本。另外,由于有标签样本很少,不符合优选准则2的有标签样本将对辅学习器的针对性产生较大影响,必须剔除以保证对无标签样本预测的准确性。通过上述分析,从数据分布的角度说明了双优选方法的必要性。

为了进一步验证方法的实际效果,分别比较了GPR(不利用无标签样本)、无优选半监督GPR(non-optimal semi-supervised GPR, NS-GPR)和本文方法(利用优选准则1和优选准则2)对真实值的跟踪效果,具体效果如图3所示。

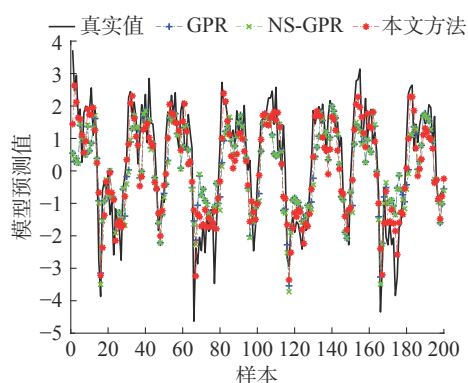


图3 数值仿真双优选半监督预测效果

Fig. 3 Numerical simulation of double-optimal semi-supervised prediction

由图3可知,本文方法对真实值的跟踪效果远远优于GPR和NS-GPR。双优选方法在充分地利用无标签样本后,对测试样本真实值的跟踪效果明显加强,拟合效果有了明显的提升。具体拟合指标如表1所示。由表中数据可知,所提

双优选方法跟踪效果相较GPR,均方根误差由1.1096下降到了0.9278,可知本文方法在有标签样本很少时也能够取得优秀的跟踪效果。

表1 数值仿真双优选的拟合性能

Table 1 The fitting performance of numerical simulation of dual-optimal selection

方法	均方根误差
GPR	1.109 6
NS-GPR	1.124 8
本文方法	0.927 8

## 4 脱丁烷塔仿真实验

为了进一步验证本文方法性能,选用脱丁烷塔过程作为对象。脱丁烷塔装置是石油炼制生产过程中脱硫与石脑油分离装置的重要组成部分<sup>[16]</sup>。该过程样本数据有7个辅助变量,分别为:塔顶温度 $x_1$ ;塔顶压力 $x_2$ ;塔顶回流量 $x_3$ ;塔顶产品流出 $x_4$ ;第六层塔板温度 $x_5$ ;塔底温度1; $x_6$ 塔底温度2; $x_7$ 塔底温度2;主导变量为塔底丁烷浓度。详细的工艺过程描述可见文献[17]。

该实验过程数据源于真实过程的实时采样,共获得2394组样本。选出150组作为有标签样本,选出500组作为无标签样本。在有标签样本中,50组作为建模样本,另外100组作为测试样本。

为了进一步分析本文算法性能,纵向比较了几点改进对模型跟踪效果的影响,具体方法如下。

1) GPR方法。不利用无标签样本,仅利用已有的有标签样本建立GPR模型,然后测试其对测试样本的跟踪效果。

2) NS-GPR方法。不利用优选准则做筛选,直接对有标签样本建模,获得辅学习器,然后预测无标签样本的标签,进而利用伪标签样本更新主学习器。

3) 第一类单优选半监督GPR(single-optimal semi-supervised GPR, SS-GPRa)。利用优选准则1筛选无标签样本,然后直接对有标签样本建模得到辅学习器,后续过程同方法2)。

4) 第2类单优选半监督GPR(简称SS-GPRb)。首先利用优选准则2筛选有标签样本,然后对其建模得到辅学习器,后续过程同方法2)。

5) 本文方法。

图4为不同方法的预测值与真实值的比较,纵坐标为测试样本真实值,横坐标为测试样本预测值,数据点离基准线越近,表示预测效果越好。

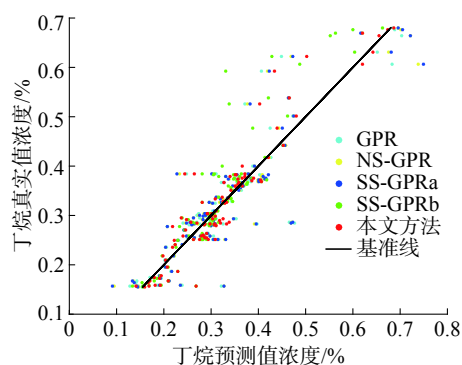


图4 不同方法的纵向比较

Fig. 4 Longitudinal comparison of different methods

分析图4发现,本文方法在几种方法中的效果最好。为了更直观地比较各方法的预测效果,图5通过跟踪误差表现了各方法的跟踪效果,其中,纵坐标为预测值与真实值的差值。

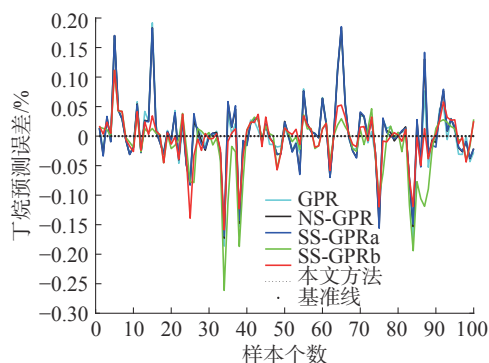
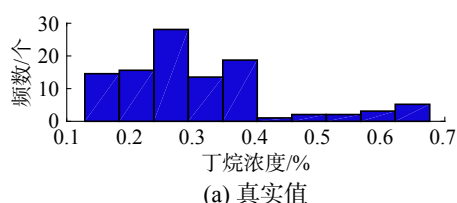
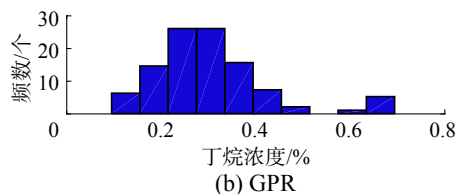


图5 不同方法的预测误差对比

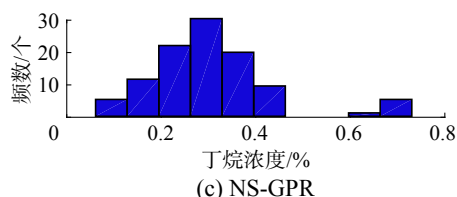
Fig. 5 Comparison of prediction errors of different methods



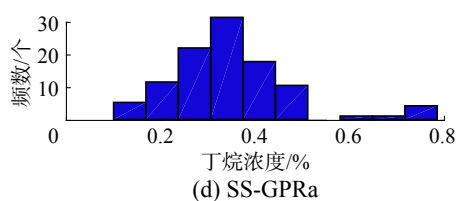
(a) 真实值



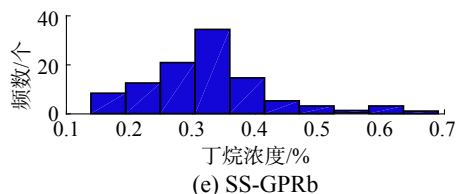
(b) GPR



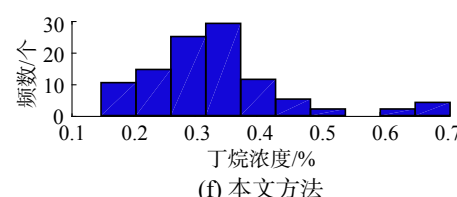
(c) NS-GPR



(d) SS-GPRa



(e) SS-GPRb



(f) 本文方法

图6 多种方法预测值与真实值的直方图统计

Fig. 6 Histogram statistics of predicted and real values of various methods

综合图4和5可以看出,NS-GPR虽然考虑了无标签样本信息,但其无差别的利用将会带来大量噪声,而SS-GPRa虽然筛选了无标签样本,但是由于辅学习器建立过程中,没有考虑有标签样本的针对性,也会带来大量噪声。而SS-GPRb在考虑筛选有标签样本后,能够对部分无标签样本进行更准确的预测,因此获得了较好的模型预测效果。

本文方法在综合考虑无标签样本与有标签样本筛选后,在绝大多数测试样本的跟踪上,都有很好的效果。这是因为在利用双优选准则后,既降低引入噪声的可能性,又保证了辅学习器的针对性,因此方法的跟踪效果有了明显的提升。

上述分析均建立在各方法对真实值的跟踪效果对比上,为了更客观地对比方法之间的效果,进行了各方法预测值的直方图统计。图6分别统计了真实值与各方法预测值的分布,其中,纵坐标为分布在每一区间的样本出现的频率,横坐标为预测值与真实值分布的区间。

由图6(a)、(c)和(d)的对比可知,NS-GPR与SS-GPRa在0.1~0.15和0.5~0.6两个区域的预测值有比较明显的误差。同样,由图6(a)与图6(b)的对比发现,GPR在0.4~0.6区域的预测值的误差也较大。进一步由图6(a)与图6(e)的对比发现,SS-GPRb在0.6~0.7区域中无法实现良好跟踪。最后,综合对比图6各方法发现,本文方法在上述区域均能进行更准确地跟踪。

上述分析对比,验证了本文方法的实际效果。发现同时考虑无标签样本与有标签样本的筛选,获得了更好的预测效果。各方法具体跟踪效果如表2所示。

表2 不同模型预测效果对比

Table 2 Comparison of prediction effects of different models

方法	均方根误差
GPR	0.055 1
NS-GPR	0.055 6
SS-GPRa	0.056 6
SS-GPRb	0.054 0
本文方法	0.040 5

## 5 结束语

本文提出的带双优选过程的半监督回归算法,基于两种优选准则,一方面筛选合适的无标签样本,另一方面筛选有标签样本,从而建立更有针对性的辅学习器,实现了对无标签样本的更准确地利用,达到了提升主学习器性能的目的。将算法进行数值仿真并应用于脱丁烷塔过程,实验结果表明,所提方法在有标签样本较少时,具有良好的预测效果,为半监督回归提供了一种新思路。

## 参考文献:

- [1] 周志华. 基于分歧的半监督学习[J]. 自动化学报, 2013, 39(11): 1871–1878.  
ZHOU Zhihua. Disagreement-based semi-supervised learning[J]. Acta Automatica Sinica, 2013, 39(11): 1871–1878.
- [2] ZHOU Zhihua, LI Ming. Semisupervised regression with cotraining-style algorithms[J]. IEEE Transactions on knowledge and data engineering, 2007, 19(11): 1479–1493.
- [3] 姜婷, 袁肖明, 岳厚光. 基于分布先验的半监督 FCM 的肺结节分类[J]. 智能系统学报, 2017, 12(5): 729–734.  
JIANG Ting, XI Xiaoming, YUE Houguang. Classification of pulmonary nodules by semi-supervised FCM based on prior distribution[J]. CAAI transactions on intelligent systems, 2017, 12(5): 729–734.
- [4] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法[J]. 计算机学报, 2015, 38(8): 1592–1617.  
LIU Jianwei, LIU Yuan, LUO Xionglin. Semi-supervised learning methods[J]. Chinese journal of computers, 2015, 38(8): 1592–1617.
- [5] 刘杨磊, 梁吉业, 高嘉伟, 等. 基于 Tri-training 的半监督多标记学习算法[J]. 智能系统学报, 2013, 8(5): 439–445.  
LIU Yanglei, LIANG Jiye, GAO Jiawei, et al. Semi-supervised multi-label learning algorithm based on Tri-training[J]. CAAI Transactions on intelligent systems, 2013, 8(5): 439–445.
- [6] 徐蓉, 姜峰, 姚鸿勋. 流形学习概述[J]. 智能系统学报, 2006, 1(1): 44–51.  
XU Rong, JIANG Feng, YAO Hongxun. Overview of manifold learning[J]. CAAI transactions on intelligent systems, 2006, 1(1): 44–51.
- [7] 杨剑, 王珏, 钟宁. 流形上的 Laplacian 半监督回归[J]. 计算机研究与发展, 2007, 44(7): 1121–1127.  
YANG Jian, WANG Jue, ZHONG Ning. Laplacian semi-supervised regression on a manifold[J]. Journal of computer research and development, 2007, 44(7): 1121–1127.
- [8] ZHOU Zhihua, LI Ming. Semi-supervised regression with co-training[C]//Proceedings of the 19th International Joint Conference on Artificial Intelligence. Edinburgh, Scotland, UK, 2005: 908–913.
- [9] 程玉虎, 冀杰, 王雪松. 基于 Help-Training 的半监督支持向量回归[J]. 控制与决策, 2012, 27(2): 205–210, 226.  
CHENG Yuhu, JI Jie, WANG Xuesong. Semi-supervised support vector regression based on Help-Training[J]. Control and decision, 2012, 27(2): 205–210, 226.
- [10] 盛高斌, 姚明海. 基于半监督回归的选择性集成算法[J]. 计算机仿真, 2009, 26(10): 198–201, 318.  
SHENG Gaobin, YAO Minghai. An ensemble selection algorithm based on semi-supervised regression[J]. Computer simulation, 2009, 26(10): 198–201, 318.
- [11] 何志昆, 刘光斌, 赵曦晶, 等. 高斯过程回归方法综述[J]. 控制与决策, 2013, 28(8): 1121–1129, 1137.  
HE Zhikun, LIU Guangbin, ZHAO Xijing, et al. Overview of Gaussian process regression[J]. Control and decision, 2013, 28(8): 1121–1129, 1137.
- [12] 熊伟丽, 李妍君, 姚乐, 等. 一种动态校正的 AGMM-GPR 多模型软测量建模方法[J]. 大连理工大学学报, 2016, 56(1): 77–85.  
XIONG Weili, LI Yanjun, YAO Le, et al. A dynamically corrected AGMM-GPR multi-model soft sensor modeling method[J]. Journal of Dalian University of Technology, 2016, 56(1): 77–85.
- [13] 郭帅, 马书根, 李斌, 等. VorSLAM 算法中基于多规则的数据关联方法[J]. 自动化学报, 2013, 39(6): 883–894.  
GUO Shuai, MA Shugen, LI Bin, et al. A data association approach based on multi-rules in VorSLAM[J]. Acta automatica sinica, 2013, 39(6): 883–894.
- [14] KNORR E M, NG R T. A unified notion of outliers: properties and computation[C]//Proceedings of the 3rd International Conference on Knowledge Discovery and Data

Mining. Newport Beach, CA, USA, 1997: 219–222.

- [15] 曾静, 王军, 郭金玉. 基于向量相似度的多模型局部建模方法研究 [J]. 计算机应用研究, 2012, 29(5): 1631–1633, 1640.

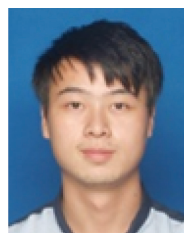
ZENG Jing, WANG Jun, GUO Jinyu. Local multi-model method based on similarity of vector[J]. Application research of computers, 2012, 29(5): 1631–1633, 1640.

- [16] 阮宏镁, 田学民, 王平. 基于联合互信息的动态软测量方法 [J]. 化工学报, 2014, 65(11): 4497–4502.

RUAN Hongmei, TIAN Xuemin, WANG Ping. Dynamic soft sensor method based on joint mutual information[J]. CIESC journal, 2014, 65(11): 4497–4502.

- [17] FORTUNA L, GRAZIANI S, RIZZO A, et al. Soft sensors for monitoring and control of industrial processes[M]. London: Springer, 2007: 229–231.

#### 作者简介:



程康明, 男, 1993 年生, 硕士研究生, 主要研究方向为工业过程建模。



熊伟丽, 女, 1978 年生, 教授, 博士, 主要研究方向为复杂工业过程建模及优化、智能优化算法及应用。主持国家自然科学基金面上项目、江苏省产学研等纵向项目 8 项; 参与国家 863 计划、重点研发计划等多项。发表研究学术论文 60 余篇。

## 2019 第二届算法、计算和人工智能国际会议 (ACAI 2019) 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2019)

ACAI 2019 将于 2019 年 12 月 20 日至 22 日在中国三亚召开, 本会议主要围绕“算法、计算和人工智能”的最新研究领域而展开, 致力于促进世界顶尖创新者、科学家、学者、研究人员和思想领导者之间的交流和探讨, 促进数据科学与信息技术领域的发展, 在会议的这三天里, 您将有机会聆听到前沿的学术报告, 见证该领域的成果与进步。

ACAI 2018 已于 2018 年成功在中国三亚召开, 会议论文集由 ACM 出版, 并已成功被 EI 和 Scopus 检索 (<http://www.acai2019.net/history.html>)。

#### 出版与检索:

所有被接收的文章将被收录在 ACAI 2019 会议论文集中, 并提交 Ei Compendex、SCOPUS、DBLP and Semantic Scholar 等数据库检索。优秀论文可推荐至国际期刊上发表。

#### 征文投递方式:

1. 通过 CMT 投稿系统提交 PDF 版本: <https://cmt3.research.microsoft.com/ACAI2019>

2. 直接将您的文章或摘要投到我们的会议邮箱, 我们收到后会第一时间回复您。

投稿邮箱: [acai@aiase.net](mailto:acai@aiase.net)

接受/拒稿通知: 论文投稿后 1~2 周。

#### 联系方式:

会议官网: <http://www.acai2019.net/>

邮箱: [acai@aiase.net](mailto:acai@aiase.net)

QQ 咨询: 3268368942

电话: +852 53465620