

(19)中华人民共和国国家知识产权局



(12)发明专利

(10)授权公告号 CN 104517162 B

(45)授权公告日 2017.09.05

(21)申请号 201410843307.8

G06Q 50/04(2012.01)

(22)申请日 2014.12.30

(56)对比文件

CN 102033523 A, 2011.04.27,

CN 101899563 A, 2010.12.01,

CN 102564644 A, 2012.07.11,

WO 2008099457 X, 2010.05.27,

彭俊 等.基于平均轨迹的连续退火过程带钢硬度预测方法.《控制工程》.2013,第20卷(第5期),

王超 等.基于正交信号校正的连续退火机组带钢质量预测方法.《仪表技术》.2012,

审查员 涂丹辉

(65)同一申请的已公布的文献号

申请公布号 CN 104517162 A

(43)申请公布日 2015.04.15

(73)专利权人 东北大学

地址 110819 辽宁省沈阳市和平区文化路3号巷11号

(72)发明人 唐立新 王显鹏

(74)专利代理机构 沈阳东大知识产权代理有限公司 21109

代理人 崔兰蔚

(51)Int.Cl.

G06Q 10/04(2012.01)

G06F 17/30(2006.01)

权利要求书2页 说明书9页 附图4页

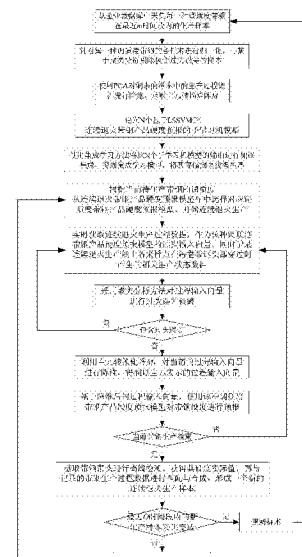
(54)发明名称

一种连续退火产品硬度在线集成学习预报方法

(57)摘要

一种连续退火产品硬度在线集成学习预报方法，属于钢铁企业连续退火生产过程的自动控制技术领域。利用企业的历史连续退火生产数据样本，使用以LSSVM作为子学习机的集成学习建模方法，针对不同调质度带钢分别建立其离线产品硬度预报模型；实际生产中，实时读取连续退火生产过程数据，并使用通过集成学习建立的离线产品硬度预报模型对当前的带钢产品硬度进行实时预报；经过实际生产数据的检验，本发明的方法能够明显提高连续退火产品硬度预报结果的精度和鲁棒性，使得现场操作人员能够实时掌握当前带钢产品的质量，并根据情况进行适时调整，弥补了离线检测大滞后的不足，从而帮助连续退火生产线提高产品质量、改进生产操作水平、增加经济效益。

B N 104517162



1. 一种连续退火产品硬度在线集成学习预报方法,其特征在于:包括如下步骤:

步骤1:针对每一种调质度带钢,从数据库中读取最近m时间段内的连续退火生产样本,得到含有B个样本的连续退火生产样本集合;

步骤2:针对每一种调质度带钢的连续退火生产样本集合,进行数据处理;

步骤3:根据步骤2数据处理后的样本集合,求取主元转换矩阵M及针对每种调质度带钢采用基于LSSVM的集成学习建模方法建立其产品硬度预报模型并形成离线连续退火带钢产品硬度预报模型库;

步骤4:根据待连续退火生产的带钢调质度,从连续退火带钢产品硬度预报模型库中选择对应调质度带钢产品硬度预报模型,开始连续退火生产;

步骤5:以连续退火生产线上采样点的最大采样周期T_{max}的z倍作为带钢产品硬度预报模型的采样周期,实时获取连续退火生产过程数据,作为该种调质度带钢产品硬度预报模型的过程输入向量,同时对连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据进行记录;所述生产过程数据为连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据与相应的带钢信息的合成;所述带钢信息包括其自身的规格信息和热轧信息;

步骤6:采用聚类分析方法对过程输入向量进行过失误差侦破,如果该过程输入向量包含过失误差,则删除该过程输入向量,放弃本周期的带钢产品硬度预报,并等待下一个采样周期的过程输入向量;否则,则进入到步骤7;

步骤7:利用步骤3中所得到的主元转换矩阵M,对当前的过程输入向量进行降维,得到以主元表示的过程输入向量;

步骤8:基于降维后的过程输入向量,使用该种调质度带钢产品硬度预报模型对带钢产品硬度进行预报;

步骤9:当本卷带钢连续退火生产完成后,截取带钢带头部分进行离线检测,获得本卷带钢产品硬度实际值,并进行记录;

步骤10:将所记录的每卷带钢的连续退火生产过程数据与其对应的产品硬度实际值进行匹配与合成,形成一个新的连续退火生产样本;每个连续退火生产样本包括H个数据项;所述H个数据项中包括若干个相关连续退火生产状态数据项、若干个带钢信息数据项和带钢产品硬度实际值;

步骤11:更新连续退火生产样本;

步骤12:按照步骤1至步骤3的方法,根据更新后的连续退火生产样本,分别对每一调质度带钢产品硬度预报模型进行更新;

步骤13:在新的带钢连续退火生产开始时,重复执行步骤4至步骤13,实现已知调质度带钢产品硬度的在线预报。

2. 根据权利要求1所述的连续退火产品硬度在线集成学习预报方法,其特征在于:所述的步骤2中针对每一种调质度带钢的连续退火生产样本集合进行数据处理,包括如下步骤:

步骤2.1:分别对样本集合内每一数据项进行归一化处理,以消除不同数据项之间量纲不同的影响;

步骤2.2:针对归一化处理后的样本集合采用基于聚类的过失误差侦破方法,剔除样本集合中A个包含过失误差的样本;

步骤2.2.1: 分别计算每个样本与其它B-1个样本之间的最小欧氏距离 d_i ($i=1, 2, \dots, B$) ;

步骤2.2.2: 计算所有样本的最小欧氏距离的平均值 d_{avg} , 将 d_{avg} 的q倍作为第一门槛;

步骤2.2.3: 将样本集合中与其它B-1个样本之间的最小欧氏距离 d_i 大于第一门槛值的样本剔除, 同时将剩余的样本进行聚类, 得到P个聚类;

步骤2.2.4: 分别计算每一聚类与其它P-1个聚类之间的最小欧氏距离 D_u ($u=1, 2, \dots, P$) ;

步骤2.2.5: 计算所有聚类的最小欧氏距离的平均值 D_{avg} , 并将 D_{avg} 的r倍作为第二门槛;

步骤2.2.6: 将与其它P-1个聚类之间的最小欧氏距离大于第二门槛值并且聚类内样本数量小于t的聚类及其包含的样本剔除。

3. 根据权利要求1所述的连续退火产品硬度在线集成学习预报方法, 其特征在于: 所述的步骤3中根据步骤2数据处理后的样本集合求取主元转换矩阵M的方法为: 利用PCA方法, 对步骤2数据处理后的样本集合中所有生产过程数据构成的矩阵进行降维, 保留特征根大于等于1的主元, 并将保留的主元存储于主元转换矩阵M。

4. 根据权利要求2所述的连续退火产品硬度在线集成学习预报方法, 其特征在于: 所述的步骤3中根据步骤2数据处理后的样本集合针对每种调质度带钢采用基于LSSVM的集成学习建模方法建立其产品硬度预报模型的方法, 包括如下步骤:

步骤3.2.1: 为步骤2数据处理后的样本集合中所有样本均赋予相同的权重值, 即每个样本的权重 $w_i = 1 / (B-A)$;

步骤3.2.2: 基于当前的样本权重, 使用LSSVM方法建立一个带钢产品硬度预报的子学习机模型, 计算该模型的均方根误差以及各样本的训练误差;

步骤3.2.3: 依据各样本的训练误差调整样本权重, 方法为: 首先使训练误差较大的样本的权重增大及训练误差较小的样本的权重减小, 得到新的样本权重, 然后对新的样本权重进行归一化处理;

步骤3.2.4: 重复执行步骤3.2.2至步骤3.2.3, 建立N个连续退火带钢产品硬度预报的子学习机模型, 并计算各子学习机模型的均方根误差以及各子学习机模型所对应的各样本的训练误差;

步骤3.2.5: 根据步骤3.2.4计算的各子学习机模型的均方根误差, 计算各子学习机模型的权重;

步骤3.2.6: 根据N个子学习机模型及其各自的权重, 将N个子学习机模型集成为一个集成学习机模型, 该集成学习机模型即为该种调质度带钢产品硬度预报模型, 其输出为各子学习机模型所输出的带钢产品硬度预报值的加权合成值;

步骤3.2.7: 将该种调质度带钢产品硬度预报模型存入离线连续退火带钢产品硬度预报模型库。

5. 根据权利要求1所述的连续退火产品硬度在线集成学习预报方法, 其特征在于: 所述的步骤11中更新连续退火生产样本的方法为: 对于某一调质度带钢, 当其最近O时间段内的连续退火生产样本采集完成后, 则将该最近O时间段内的连续退火生产样本加入到企业数据库中, 同时将该数据库中最早O时间段内的连续退火生产样本删除。

一种连续退火产品硬度在线集成学习预报方法

技术领域

[0001] 本发明属于钢铁企业连续退火生产过程的自动控制技术领域,特别涉及一种连续退火产品硬度在线集成学习预报方法。

背景技术

[0002] 在钢铁企业冷轧厂连续退火机组的实际生产过程中,带钢的硬度是衡量产品质量与指导生产的核心指标。在实际生产过程中,带钢的硬度还无法实现在线检测,现场是通过截取退火后带钢的头、尾部分,再进行离线实验分析来测定带钢的硬度,从而判断产品质量情况。但是,通过离线实验分析来获得产品的质量信息一般都具有一定的时间滞后性,也就是说只有当带钢生产出来一段时间后,才能获得其具体的质量信息,而实际生产中带钢的生产速度非常快,在几分钟之内就能完成退火处理,从而经常发生带钢硬度波动非常大的情况,导致产生硬度不达标甚至废品等质量问题,严重影响冷轧厂的经济效益。

[0003] 论文“基于PLS的连续退火机组带钢质量预报及过程监测系统设计与实现[D]”(汪源,东北大学,2009)虽然针对带钢产品硬度提出了一种基于偏最小二乘(Partial Least Squares,PLS)的数据驱动型建模方法,但是这篇论文中所提出的方法并不能满足实际生产过程的需要,主要原因有:(1)该文献所考虑的与带钢硬度相关的信息较少,仅有20个左右,而实际生产过程中与带钢硬度相关的过程信息则要多达51个;(2)该文献所提出的PLS方法主要还是针对连续退火生产过程的监测与故障诊断,并且PLS方法属于一种线性回归方法,而实际的生产过程则是非线性的,从而导致PLS方法的精度不高;(3)样本数据存在输入项众多并且相互耦合、而输出结果范围却较窄的问题,导致样本之间存在输入项差别较大而输出结果却相同或相近,使得传统的数据建模方法难以获得较高的预测精度和鲁棒性。

发明内容

[0004] 针对现有技术存在的不足,本发明提供一种连续退火产品硬度在线集成学习预报方法。

[0005] 本发明的技术方案是:

[0006] 一种连续退火产品硬度在线集成学习预报方法,包括如下步骤:

[0007] 步骤1:针对每一种调质度带钢,从数据库中读取最近m时间段内的连续退火生产样本,得到含有B个样本的连续退火生产样本集合;

[0008] 步骤2:针对每一种调质度带钢的连续退火生产样本集合,进行数据处理;

[0009] 步骤2.1分别对样本集合内每一数据项进行归一化处理,以消除不同数据项之间量纲不同的影响;

[0010] 步骤2.2:针对归一化处理后的样本集合采用基于聚类的过失误差侦破方法,剔除样本集合中A个包含过失误差的样本;

[0011] 步骤2.2.1:分别计算每个样本与其它B-1个样本之间的最小欧氏距离 d_i ($i = 1, 2, \dots, B$) ;

- [0012] 步骤2.2.2:计算所有样本的最小欧氏距离的平均值 d_{avg} ,将 d_{avg} 的q倍作为第一门槛;
- [0013] 步骤2.2.3:将样本集合中与其它B-1个样本之间的最小欧氏距离 d_i 大于第一门槛值的样本剔除,同时将剩余的样本进行聚类,得到P个聚类;
- [0014] 步骤2.2.4:分别计算每一聚类与其它P-1个聚类之间的最小欧氏距离 D_u ($u=1, 2, \dots, P$);
- [0015] 步骤2.2.5:计算所有聚类的最小欧氏距离的平均值 D_{avg} ,并将 D_{avg} 的r倍作为第二门槛;
- [0016] 步骤2.2.6:将与其它P-1个聚类之间的最小欧氏距离大于第二门槛值并且聚类内样本数量小于t的聚类及其包含的样本剔除;
- [0017] 步骤3:根据步骤2数据处理后的样本集合,求取主元转换矩阵M及针对每种调质度带钢采用基于LSSVM的集成学习建模方法建立其产品硬度预报模型并形成离线连续退火带钢产品硬度预报模型库;
- [0018] 步骤3.1:利用PCA方法,对步骤2数据处理后的样本集合中所有生产过程数据构成的矩阵进行降维,保留特征根大于等于1的主元,并将保留的主元存储于主元转换矩阵M;
- [0019] 步骤3.2:根据步骤2数据处理后的样本集合,采用基于LSSVM的集成学习建模方法,针对每种调质度带钢建立其产品硬度预报模型,形成离线连续退火带钢产品硬度预报模型库;所述针对每种调质度带钢建立其产品硬度预报模型的方法,包括如下步骤:
- [0020] 步骤3.2.1:为步骤2数据处理后的样本集合中所有样本均赋予相同的权重值,即每个样本的权重 $w_i = 1/(B-A)$;
- [0021] 步骤3.2.2:基于当前的样本权重,使用LSSVM方法建立一个带钢产品硬度预报的子学习机模型,计算该模型的均方根误差以及各样本的训练误差;
- [0022] 步骤3.2.3:依据各样本的训练误差调整样本权重,方法为:首先使训练误差较大的样本的权重增大及训练误差较小的样本的权重减小,得到新的样本权重,然后对新的样本权重进行归一化处理;
- [0023] 步骤3.2.4:重复执行步骤3.2.2至步骤3.2.3,建立N个连续退火带钢产品硬度预报的子学习机模型,并计算各子学习机模型的均方根误差以及各子学习机模型所对应的各样本的训练误差;
- [0024] 步骤3.2.5:根据步骤3.2.4计算的各子学习机模型的均方根误差,计算各子学习机模型的权重;
- [0025] 步骤3.2.6:根据N个子学习机模型及其各自的权重,将N个子学习机模型集成为一个集成学习机模型,该集成学习机模型即为该种调质度带钢产品硬度预报模型,其输出为各子学习机模型所输出的带钢产品硬度预报值的加权合成值;
- [0026] 步骤3.2.7:将该种调质度带钢产品硬度预报模型存入离线连续退火带钢产品硬度预报模型库;
- [0027] 步骤4:根据待连续退火生产的带钢调质度,从连续退火带钢产品硬度预报模型库中选择对应调质度带钢产品硬度预报模型,开始连续退火生产;
- [0028] 步骤5:以连续退火生产线上采样点的最大采样周期 T_{max} 的z倍作为带钢产品硬度预报模型的采样周期,实时获取连续退火生产过程数据,作为该种调质度带钢产品硬度预

报模型的过程输入向量,同时对连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据进行记录;所述生产过程数据为连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据与相应的带钢信息的合成;所述带钢信息包括其自身的规格信息和热轧信息(带钢在热轧工序的生产信息);

[0029] 步骤6:采用聚类分析方法对过程输入向量进行过失误差侦破,如果该过程输入向量包含过失误差,则删除该过程输入向量,放弃本周期的带钢产品硬度预报,并等待下一个采样周期的过程输入向量;否则,则进入到步骤7;

[0030] 步骤7:利用步骤3中所得到的主元转换矩阵M,对当前的过程输入向量进行降维,得到以主元表示的过程输入向量;

[0031] 步骤8:基于降维后的过程输入向量,使用该种调质度带钢产品硬度预报模型对带钢产品硬度进行预报;

[0032] 步骤9:当本卷带钢连续退火生产完成后,截取带钢带头部分进行离线检测,获得本卷带钢产品硬度实际值,并进行记录;

[0033] 步骤10:将所记录的每卷带钢的连续退火生产过程数据与其对应的产品硬度实际值进行匹配与合成,形成一个新的连续退火生产样本;每个连续退火生产样本包括H个数据项;所述H个数据项中包括若干个相关连续退火生产状态数据项、若干个带钢信息数据项和带钢产品硬度实际值;

[0034] 步骤11:更新连续退火生产样本,方法为:对于某一调质度带钢,当其最近0时间段内的连续退火生产样本采集完成后,则将该最近0时间段内的连续退火生产样本加入到企业数据库中,同时将该数据库中最早0时间段内的连续退火生产样本删除;

[0035] 步骤12:按照步骤1至步骤3的方法,根据更新后的连续退火生产样本,分别对每一调质度带钢产品硬度预报模型进行重新训练和更新;

[0036] 步骤13:在新的带钢连续退火生产开始时,重复执行步骤4至步骤13,实现已知调质度带钢产品硬度的在线预报。

[0037] 本发明的有益效果:本发明的连续退火产品硬度在线集成学习预报方法使用LSSVM作为子学习机生成器,并使用集成学习框架将各子学习机的结果集成起来,作为最终的带钢产品硬度预报值。经过实际生产数据的检验,本发明的方法能够明显提高连续退火产品硬度预报结果的精度和鲁棒性,使得现场操作人员能够实时掌握当前带钢产品的质量,并根据情况进行适时调整,弥补了离线检测大滞后的不足,从而帮助连续退火生产线提高产品质量、改进生产操作水平、增加经济效益。

附图说明

[0038] 图1为本发明一种实施方式的连续退火产品硬度在线集成学习预报方法原理示意图;

[0039] 图2为本发明一种实施方式的连续退火产品硬度在线集成学习预报方法流程图;

[0040] 图3为本发明一种实施方式的连续退火产品硬度在线集成学习预报方法的预报界面图;

[0041] 图4为本发明一种实施方式的连续退火产品硬度集成学习预报值与相应测试样本的比较结果图;

[0042] 图5为本发明一种实施方式的连续退火产品硬度在线监测界面图；

具体实施方式

[0043] 下面结合附图和具体实施方式对本发明作进一步说明。

[0044] 连续退火生产是钢铁企业冷轧厂的一道重要工序，如图1所示，连续退火生产线按照功能可以分为以下9个阶段：加热炉(HF)、均热炉(SF)、缓冷炉(SCF)、1#冷炉(1C)、1#过时效炉(10A)、2#过时效炉(20A)、2#冷却炉(2C)、水淬炉(WQ)、平整机。在生产过程中，冷轧带钢以一定的速度在生产线各炉内穿行，使其按照设定的退火工艺路线完成加热、冷却等热处理工艺过程，从而消除带钢因冷轧所造成的内应力，再经过平整后，得到高质量的带钢。

[0045] 针对在实际生产过程中，带钢的硬度还无法实现在线检测的现状，本发明提供一种连续退火产品硬度在线集成学习预报方法。该方法的工作原理如图1所示：首先，利用企业的历史连续退火生产数据样本，使用以LSSVM(Least Square Support Vector Machine，最小二乘支持向量机)作为子学习机的集成学习建模方法，针对不同调质度带钢分别建立其离线产品硬度预报模型，并形成离线连续退火带钢产品硬度预报模型库；其次，在实际生产过程中，根据待连续退火生产的带钢调质度从离线建立的连续退火带钢产品硬度预报模型库中选择相应的产品硬度预报模型；然后，从连续退火生产过程采样与控制系统中获取连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据并进行数据预处理且与带钢信息进行合成得到连续退火实时生产过程数据；接着，使用聚类分析方法对连续退火实时生产过程数据进行过失误差侦破；对于不包含过失误差的连续退火实时生产过程数据，经过PCA(Principal Component Analysis, 主元分析)方法降维处理后，传递给所选择的产品硬度预报模型，从而向现场操作人员输出当前工况下的带钢产品硬度预报值，进而便于现场操作人员及时掌握产品质量并根据当前的带钢产品硬度预报值进行相应的操作条件调整。此外，每卷带钢带头处的生产过程数据会存储起来，当带钢产品硬度离线检测结果出来以后，与其合成后，形成新的连续退火生产数据样本，存储到企业数据库中。当最近一个月的新样本采集完成后，会启动带钢产品硬度预报模型的重新训练过程，从而保证预报模型能够跟踪连续退火机组的最新工况。

[0046] 本实施方式基于图1所示的连续退火生产工艺，生产调质度为T5的带钢产品，并根据连续退火产品硬度在线集成学习预报方法原理，将本发明所提出的连续退火产品硬度在线集成学习预报方法应用于本实施方式，如图2所示，包括如下步骤：

[0047] S1：针对调质度为T5的带钢，从企业数据库中采集最近3个月的连续退火生产样本，共计获得350个样本；其中每个样本由以下52个数据项构成，其中包括14个带钢信息数据项，分别为9个带钢规格信息数据项即入口厚度、入口宽度、碳含量、硫含量、磷含量、锰含量、氮含量、硅含量和总铝含量，以及5个热轧信息数据项即出炉温度、平均卷曲温度、平均精轧温度、平均终轧温度和连续酸洗-轧制延伸率(CDCM延伸率)；37个连续退火相关生产状态数据项，分别为中央段速度、HF炉1区炉温、HF炉2区炉温、HF炉3区炉温、HF炉4区炉温、HF炉5区炉温、HF炉出口带钢温度、SF炉1区炉温、SF炉2区炉温、SF炉出口带钢温度、SCF炉1区炉温、SCF炉2区炉温、SCF炉出口带钢温度、1C炉1区炉温、1C炉2区炉温、1C炉3区炉温、1C炉冷却气体温度、1C炉出口带钢温度、10A炉1区炉温、10A炉2区炉温、10A炉出口带钢温度、20A炉1-1区炉温、20A炉1-2区炉温、20A炉2-1区炉温、20A炉2-2区炉温、20A炉出口带钢温度、2C

炉炉温、2C冷却气体温度、2C炉出口带钢温度、WQ炉1区水温、WQ炉2区水温、平整机延伸率、平整机入口带钢张力、平整机中间带钢张力、平整机出口带钢张力、平整机1#机架轧制力和2#机架轧制力；1个带钢产品硬度实际值数据项。所采集的连续退火生产样本集合如表1所示，其中每一行表示一个样本，每一列作为一个数据项，并依次按照从带钢规格信息数据项、热轧信息数据项、连续退火相关生产状态数据项至最后的带钢产品硬度实际值数据项的顺序进行排列。以 S_i 表示样本， $S_i = (s_{i1}, s_{i2}, \dots, s_{ik}, \dots, s_{i52})$ ，其中 $i=1, 2, \dots, 350$ ， s_{ik} 表示第*i*个样本的第*k*个数据项。

[0048] 表1.最近3个月内调质度为T5的带钢产品的连续退火生产样本集合

[0049]

编号	入口厚度 (mm)	入口宽度 (mm)	碳含量 (%)	硫含量 (%)	磷含量 (%)	...	1#平整机轧制力(kN)	2#平整机轧制力(kN)	硬度
1	213	9070	1.010	0.0043	0.0012	...	969.89	489.84	65
2	213	9070	1.010	0.0043	0.0012	...	965.80	488.12	66
3	170	8560	0.968	0.0060	0.0013	...	965.87	502.02	65
...
349	182	8260	1.006	0.0063	0.0099	...	961.81	407.30	62
350	219	9130	0.987	0.0039	0.0015	...	965.80	536.82	64

[0050] S2:对表1所示的连续退火生产样本集合进行数据处理；

[0051] S2.1:分别对表1中的每一列数据项进行归一化处理，以消除不同数据项之间量纲不同的影响；

[0052] S2.2:针对归一化处理后的样本集合采用基于聚类的过失误差侦破方法，剔除包含过失误差的样本，包括以下步骤：

[0053] S2.2.1: 分别计算每个样本 S_i 与其他349个样本之间的最小欧氏距离

$$d_i = \min_{j=1, \dots, 350, j \neq i} \sqrt{\sum_{k=1}^{52} (s_{ik} - s_{jk})^2}, \text{ 其中 } i=1, 2, \dots, 350;$$

[0054] S2.2.2:计算所有样本的最小欧氏距离的平均值 $d_{avg} = \sum_{i=1}^{350} d_i / 350$ ，将 d_{avg} 的3倍，即 $3d_{avg}$ 作为第一门槛值($3d_{avg}$ 为经过实验获得的门槛值)；

[0055] S2.2.3:将样本集合中与其他349个样本的最小欧氏距离为 $d_i > 3d_{avg}$ 的样本剔除，同时将剩余的样本(与其他349样本的最小欧氏距离为 $d_i \leq 3d_{avg}$ 的样本)按照就近原则进行聚类，得到36个聚类；

[0056] S2.2.4:分别计算每一聚类 C_u ($u=1, 2, \dots, 36$) 与其它35个聚类之间的最小欧氏距

离 $D_u = \min_{v=1, \dots, 36, v \neq u} \sqrt{\sum_{k=1}^{52} (c_{uk} - c_{vk})^2}$ ，其中 $c_{uk} = \sum_{m=1}^{|C_u|} s_{mk} / |C_u|$ ，表示聚类 C_u 的中心 (c_{u1}, \dots, c_{u52}) 中第*k*个数据项的值，即该聚类所包含的所有样本中第*k*个数据项的算术平均值，其中 $|C_u|$ 表示聚类 C_u 中样本的数量；

[0057] S2.2.5:计算36个聚类的最小欧氏距离的平均值 D_{avg} ，并将 D_{avg} 的3倍，即 $3D_{avg}$ 作为第二门槛；

[0058] S2.2.6:将与其它35个聚类之间的最小欧氏距离大于第二门槛值并且聚类内样本

数量小于10的聚类及其包含的样本剔除;其中 $3D_{avg}$ 和聚类内的样本数量10为经过实验获得的门槛值;

[0059] 表1所示的连续退火生产样本集合经过S2.1和S2.2的数据处理后,共有5个包含过失误差的样本被剔除,如表2所示。

[0060] 表2. 被剔除的包含过失误差的5个样本

[0061]

编号	入口厚度 (mm)	入口宽度 (mm)	碳含量 (%)	硫含量 (%)	磷含量 (%)	...	1#平整机轧 制力(kN)	2#平整机轧 制力(kN)	硬度
135	180	9570	1.060	0.0057	0.00129	...	915.15	438.12	64
308	282	7980	0.663	0.0088	0.00126	...	788.57	386.43	62
309	282	7980	0.663	0.0088	0.00126	...	785.42	382.66	62
310	223	9150	0.984	0.0096	0.0014	...	967.25	447.07	62
326	282	8860	0.951	0.0047	0.0013	...	838.48	437.37	63

[0062] S3: 使用PCA方法对样本集合中剩余的345个样本的前14个带钢信息数据项和37个相关生产状态数据项共51个数据项构成的矩阵进行降维,保留特征根大于等于1的11个主元,存储于主元转换矩阵M;同时,还根据样本集合中剩余的345个样本,采用基于LSSVM的集成学习建模方法,针对该种调质度带钢建立其产品硬度预报模型,并将所建立的该种调质度带钢产品硬度预报模型存储于离线连续退火带钢产品硬度预报模型库中;所述针对该种调质度带钢建立其产品硬度预报模型的方法,包括如下步骤:

[0063] S3.1: 为345个样本均赋予相同的权重值,即每个样本的权重 $w_i = 1/345$,并设置学习机索引 $i=1$;

[0064] S3.2: 基于当前的样本权重,使用LSSVM方法训练并得到一个带钢产品硬度预报的子学习机模型;方法为:在LSSVM中使用径向基核函数,即 $K(x, x_i) = e^{-\|x-x_i\|^2/\sigma^2}$, 其中 x_i 表示第*i*个训练样本的前51个输入数据项, σ^2 为核函数参数,具体的步骤如下:

[0065] S3.2.1: 在取值范围内,随机产生20个初始的LSSVM参数组合。记每个参数组合为 $P_j = (\gamma_j, \sigma_j^2)$, 其中 γ_j 与 σ_j^2 分别表示参数组合 P_j 中对误差的惩罚系数和核函数参数。 γ_j 与 σ_j^2 的取值范围分别为[5, 50]与[0.1, 2.0];

[0066] S3.2.2: 针对每一个参数组合 $P_j = (\gamma_j, \sigma_j^2)$, 求解如下所示的一个线性方程组:

$$[0067] \quad \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma_j} & \cdots & K(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n, x_1) & \cdots & K(x_n, x_n) + \frac{1}{\gamma_j} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 1 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

[0068] 其中 $n=345$ 为样本总数量, y_i ($i=1, \dots, 345$) 表示第*i*个样本的输出,即带钢硬度。从而获得该参数组合所对应的LSSVM的模型参数 α_i ($i=1, \dots, 345$) 与 b ,再根据各样本数据的前51个输入数据项使用所建立的LSSVM模型来预测其硬度,即硬度预测值

$\hat{y}_i = \sum_{j=1}^n \alpha_j K(x_i, x_j) + b$, 进而计算该参数组合 P_j 的评价指标,即预测结果的均方根误差

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}.$$

[0069] S3.2.3: 将当前RMSE值最小的LSSVM模型作为最好的LSSVM模型, 记为LSSVM_{best}。

[0070] S3.2.4: 基于当前的20个参数组合, 按照以下方法得到新的20个参数组合: 首先, 随机选择两个参数组合P_f=(γ_f, σ²_f)与P_g=(γ_g, σ²_g); 其次, 产生一个新的参数组合P_j=(γ_j, σ²_j), γ_j在范围[γ_{min}-0.5I₁, γ_{max}+0.5I₁]内随机产生, 其中γ_{min}=min{γ_f, γ_g}, γ_{max}=max{γ_f, γ_g}, I₁=γ_{max}-γ_{min}; 相似地, σ²_j在范围[σ²_{min}-0.5I₂, σ²_{max}+0.5I₂]内随机产生, 其中σ²_{min}=min{σ²_f, σ²_g}, σ²_{max}=max{σ²_f, σ²_g}, I₂=σ²_{max}-σ²_{min}。

[0071] S3.2.5: 对于新产生的20个参数组合, 使用S3.2.2所提出的方法计算其评价指标。从当前40个参数组合中取评价指标较小的前20个参数组合, 如果得到了更好的LSSVM模型, 则更新LSSVM_{best}, 删去剩余的20个质量较差的参数组合。

[0072] S3.2.6: 重复执行S3.2.4和S3.2.5, 共计500次(重复500次是为了能够在给定的运行时间范围内获得质量较好的LSSVM模型), 将所得到的LSSVM_{best}存储起来, 作为第1个子学习机, 并计算该模型所对应的各样本x_i的训练误差ζ_i=(̂y_i-y_i)/y_i;

[0073] S3.3: 依据各样本x_i的训练误差ζ_i调整样本权重, 方法为: 首先, 设置ε=0, 针对所有样本, 如果ζ_i>0.01, 则ε=ε+w_i; 其次, 设置β₁=ε×ε; 再次, 如果样本x_i的训练误差ζ_i>0.01, 则调整该样本的权重为w_i=w_i×β₁, 否则其权重调整为w_i=1; 从而使得训练误差较大的样本的权重增大, 而训练误差较小的样本权重变小; 最后, 对所有的权重w_i进行归一化处理。

[0074] S3.4: 重复使用上述S3.2和S3.3所提出的方法, 建立5个带钢产品硬度预报的子学习机模型, 各子学习机模型的最优参数组合及其相应的β₁值、平均训练误差如表3所示。

[0075] 表3. 带钢产品硬度预报子学习机模型的相关参数

[0076]

编号1	γ	σ ²	β ₁	平均相对误差(%)	RMSE
1	23.408	0.806	0.0125	0.50059	4.12081
2	20.9353	0.845	0.778	0.539676	4.34022
3	49.6097	0.731	0.570	0.533521	4.30655
4	25.0794	0.702	0.563	0.604213	4.95296
5	21.8188	0.712	0.181	0.480041	3.86035

[0077] S3.5: 基于所建立的5个子学习机模型, 依据如下公式确定集成学习机模型的带钢产品硬度预测结果: ̂y_{final}(x)=(̂y₁(x)+̂y₂(x)+̂y₃(x)+̂y₄(x)+̂y₅(x))/5, 其中̂y_i(x)是针对样本x的前51个输入项, 由第1个带钢产品硬度预报模型所计算出的预测值。5个子学习机模型集成后的T5调质度带钢产品硬度预报模型的训练结果如图3所示。

[0078] 除所使用的最近3个月的345个训练样本之外, 从企业数据库中采集时间更早的1个月的41个带钢产品硬度实际值作为T5调质度带钢产品硬度预报模型的测试样本, 以检验所提出的基于集成学习的连续退火产品硬度预报模型的有效性。通过T5调质度带钢产品硬度预报模型所得到的带钢产品硬度预测结果与这41个带钢产品硬度实际值的比较结果如

表4和图4所示,可以看出,本发明所提出的方法在连续退火产品硬度的预测性能上,相对没有学习过的测试样本的平均预测偏差为0.585%,并且能够跟随带钢产品实际硬度的变化趋势,具有较好的泛化能力和鲁棒性。

[0079] 表4. 测试样本与基于集成学习的连续退火产品硬度预报模型的预测结果对比

[0080]

编号	硬度实际 检测值	预测值	偏差(%)	编号	硬度实际 检测值	预测值	偏差(%)
1	63	63.59	0.937	21	65	64.43	0.877
2	63	63.52	0.825	22	64	64.43	0.672
3	64	63.55	0.703	23	64	64.41	0.641
4	63	63.77	1.222	24	63	63.31	0.492
5	64	63.76	0.375	25	64	64.17	0.266
6	65	64.60	0.615	26	64	63.95	0.078
7	65	64.45	0.846	27	65	64.37	0.969
8	65	65.00	0.000	28	63	63.40	0.635
9	64	64.11	0.172	29	64	63.20	1.250
10	64	64.39	0.609	30	64	63.83	0.266
11	63	63.08	0.127	31	65	64.51	0.754
12	63	63.48	0.762	32	64	63.91	0.141
13	64	64.27	0.422	33	65	64.78	0.338
14	64	63.99	0.016	34	63	63.30	0.476
15	64	63.85	0.234	35	62	62.34	0.548
16	63	63.36	0.571	36	64	63.89	0.172
17	64	64.51	0.797	37	62	62.65	1.048
18	65	64.71	0.446	38	63	64.02	1.619
19	65	64.25	1.154	39	64	63.99	0.016
20	64	63.38	0.969	40	62	61.93	0.113
				41	64	64.52	0.812

[0081] S3.6:将该调质度带钢产品硬度预报的集成学习模型存入离线连续退火带钢产品硬度预报模型库;

[0082] S4:在实际连续退火生产调质度为T5的带钢时,从连续退火带钢产品硬度预报模型库中选择上述T5调质度带钢产品硬度预报模型,开始连续退火生产;

[0083] S5:以连续退火生产线上采样点的最大采样周期为1秒,取其2倍作为T5调质度带钢产品硬度预报模型的采样周期,每隔2秒从连续退火生产过程采样与控制系统中读取各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据,并将其与带钢信息合成为连续退火生产过程数据,即连续退火生产样本的前51项,称为过程输入向量,同时,对连续退火生产线上各采样点在每卷带钢产品头部穿过时产生的相关生产状态数据进行记录;

[0084] S6:使用聚类分析方法对过程输入向量进行过失误差侦破,方法为:首先计算过程输入向量与当前数据库中T5调质度带钢的各连续退火生产样本的前51项之间的最小欧氏

距离,再计算最小欧氏距离平均值,如果该平均值大于S2.2.2中所计算出的第一门槛值 $3d_{avg}$,则认为其为正常样本,则进入到下一步骤S7中;否则认为该样本包含过失误差,删除该过程输入向量,放弃本次周期的硬度预报,并等待下一个采样周期的过程输入向量;

[0085] S7:利用步骤3中所得到的主元转换矩阵M,对当前的过程输入向量进行降维,得到以主元表示的过程输入向量;

[0086] S8:基于降维后的过程输入向量,使用T5调质度带钢产品硬度预报模型预报对应的带钢硬度值,并通过操作平台显示装置进行显示,如图5所示,该界面左上角会实时显示当前生产工况所对应的带钢硬度预报值,方便现场操作人员通过该预报值对生产过程进行监控,从而保证产品质量在合格范围内。

[0087] S9:当前带钢生产完成后,截取带钢带头部分进行离线检测,获得本卷带钢产品硬度实际值为64,并进行记录;

[0088] S10:将所记录的本卷带钢的连续退火生产过程数据与本卷带钢产品硬度实际值64进行匹配与合成,形成一个新的连续退火生产样本,并进行记录;

[0089] S11:更新T5调质度带钢连续退火生产样本,方法为:针对T5调质度带钢,如果当其最近一个月的连续退火生产样本采集完成后,则首先将这一个月的连续退火生产样本加入到企业数据库中,然后将该数据库中时间最早的那一个月的所有连续退火生产样本均删除,

[0090] S12:按照S2至S4的方法,根据更新后的T5调质度带钢连续退火生产样本,对T5调质度带钢产品硬度预报模型进行重新训练与更新,从而使得T5调质度带钢产品硬度预报模型能够跟随当前的连续退火实际生产工况进行相应的调整。

[0091] S13:在新的T5调质度带钢连续退火生产开始时,重复执行S4至S13,实现对T5调质度带钢产品硬度的在线预报。

[0092] 本实施方式通过连续退火产品硬度在线集成学习预报系统实现本实施方式的连续退火产品硬度在线集成学习预报方法,该系统包括以下模块:数据库维护模块、连续退火生产过程建模模块、数据预处理模块和产品硬度在线预报模块。数据库维护模块主要用来完成对企业历史数据的删除和新样本数据导入等操作;数据预处理模块主要用来完成针对生产过程样本的过失误差侦破和PCA降维等操作。连续退火生产过程建模模块用以完成基于LSSVM的集成学习建模功能,并显示所建立的带钢产品硬度预报模型的训练和测试结果。产品硬度在线预报模块主要用来实现上连续退火带钢产品硬度在线预报功能。本实施方式的连续退火产品硬度在线集成学习预报系统,是与图1所示的连续退火生产过程采样与控制系统相配套使用的,在现场使用时会将其安装到连续退火生产过程采样与控制系统所在的计算机中,该计算机作为本发明方法的硬件平台。本发明方法将从该计算机所安装的连续退火生产过程采样与控制系统中实时读取连续退火生产过程各采样点的相关生产状态数据,进行预处理后,形成连续退火生产过程数据,再使用集成学习模型对当前的带钢产品硬度进行实时预报,并将结果显示到本计算机的显示屏上,方便现场操作人员实时掌握带钢的质量信息,并在质量波动时采取相关的调整措施。

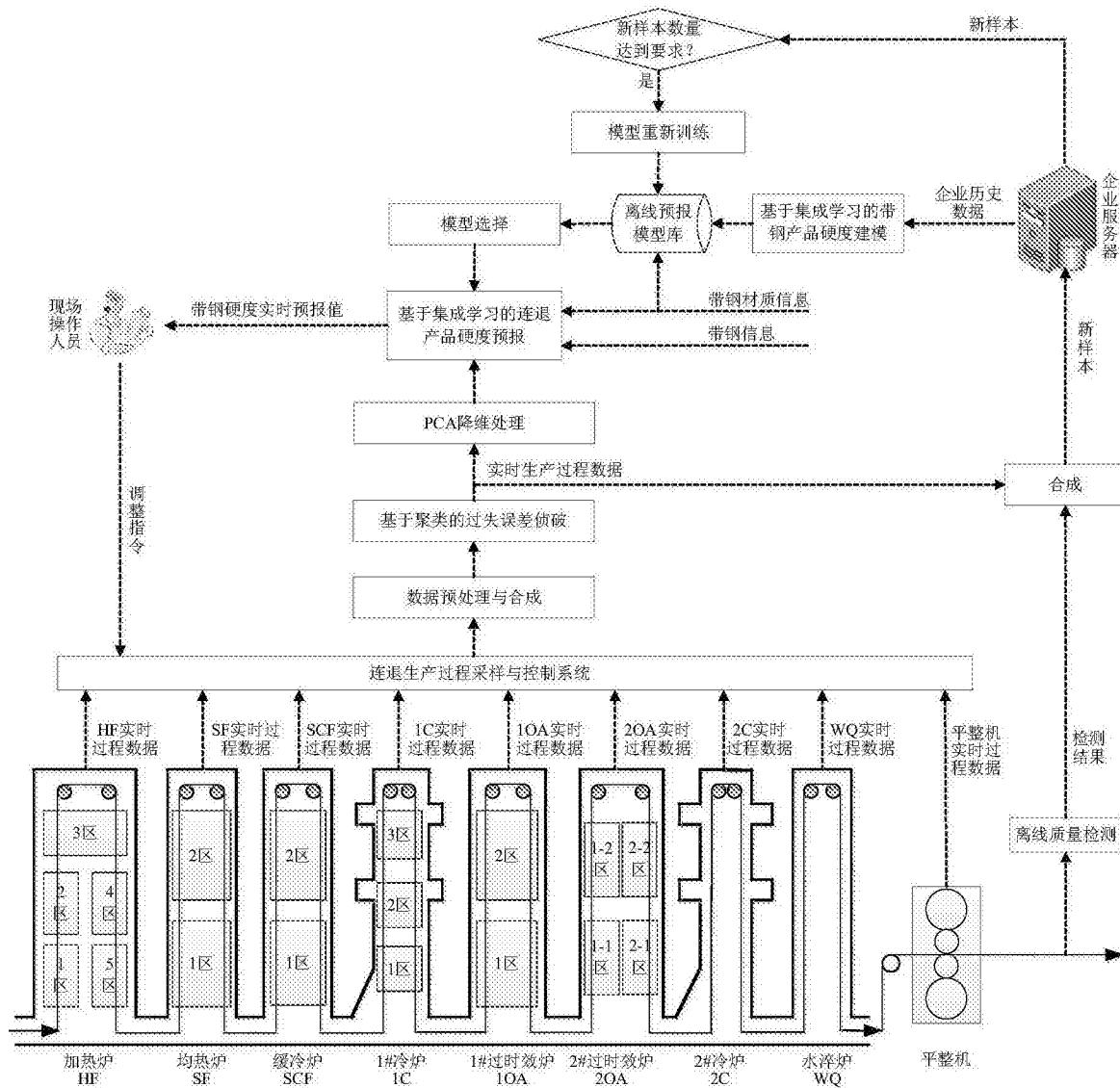


图1

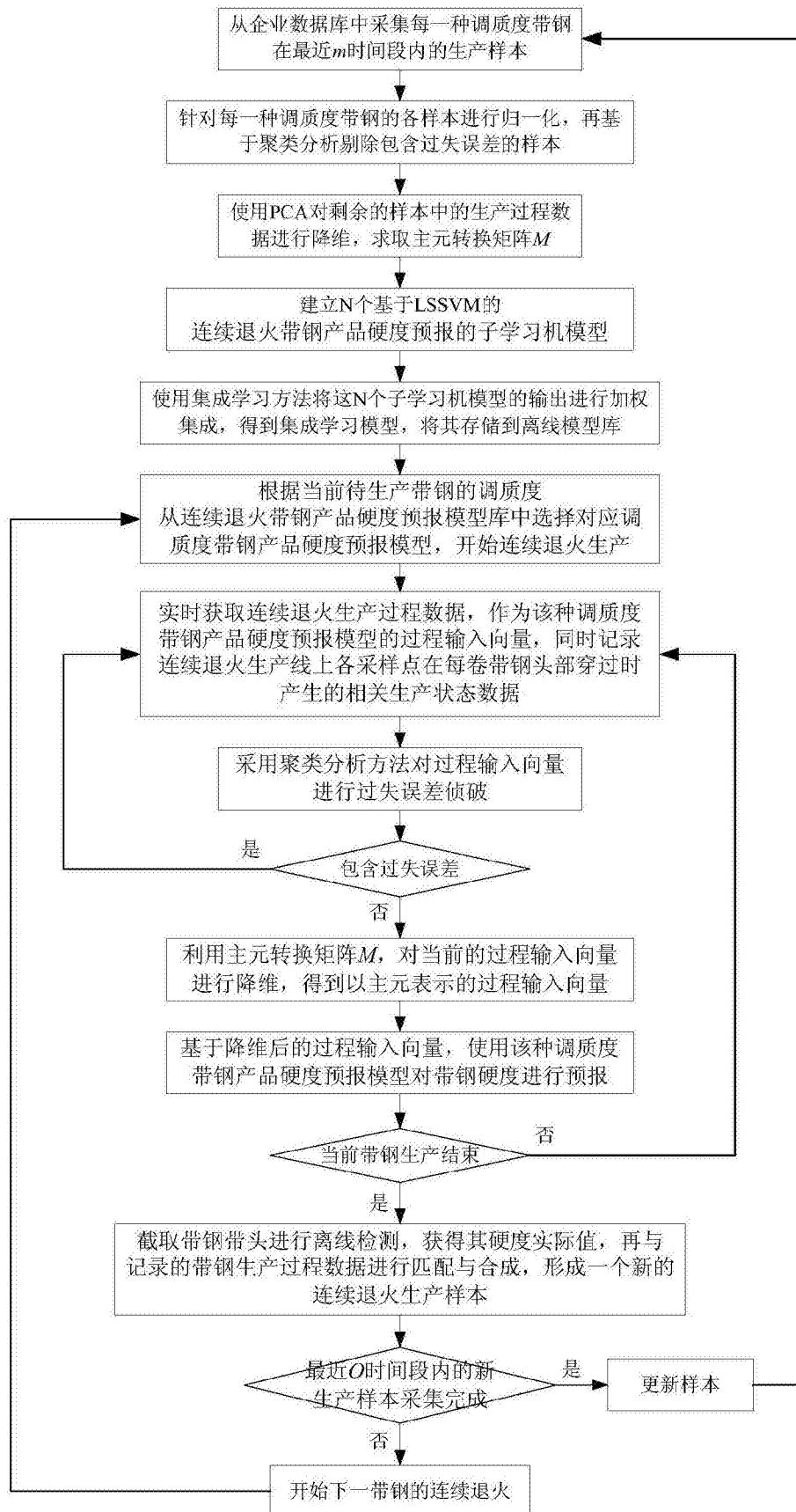


图2

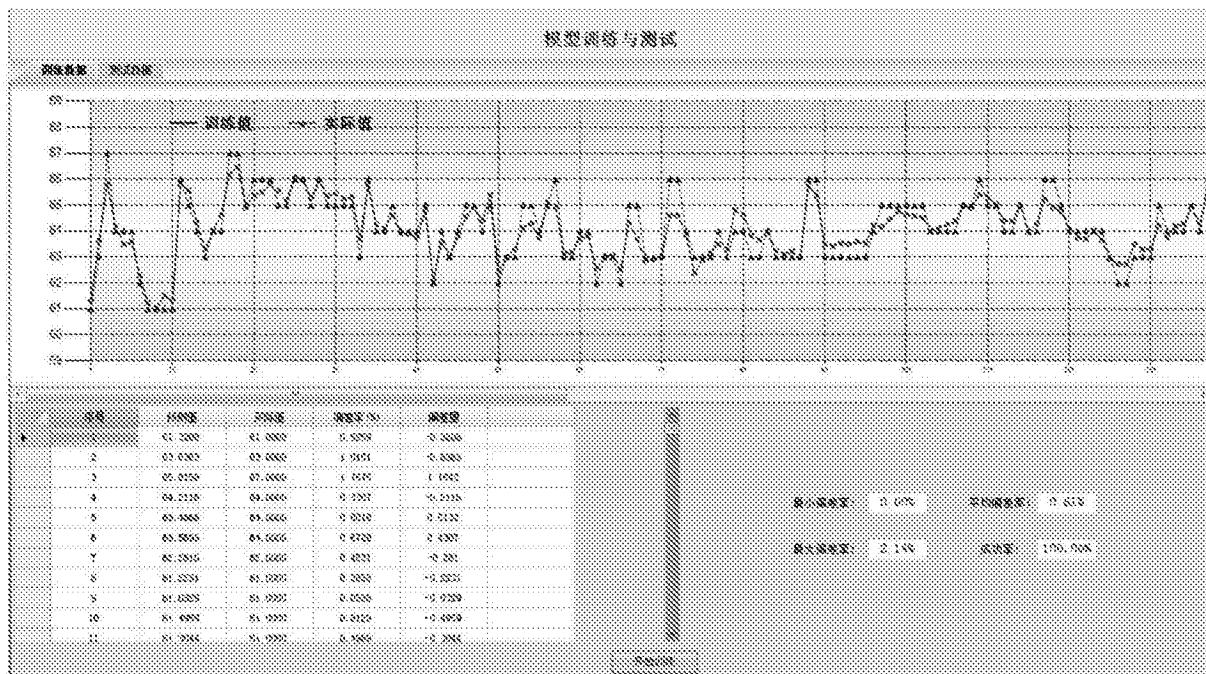


图3

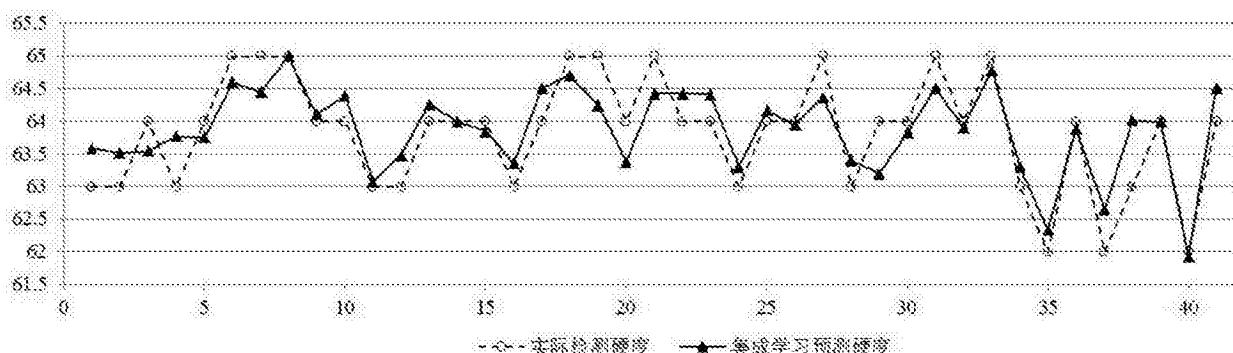


图4

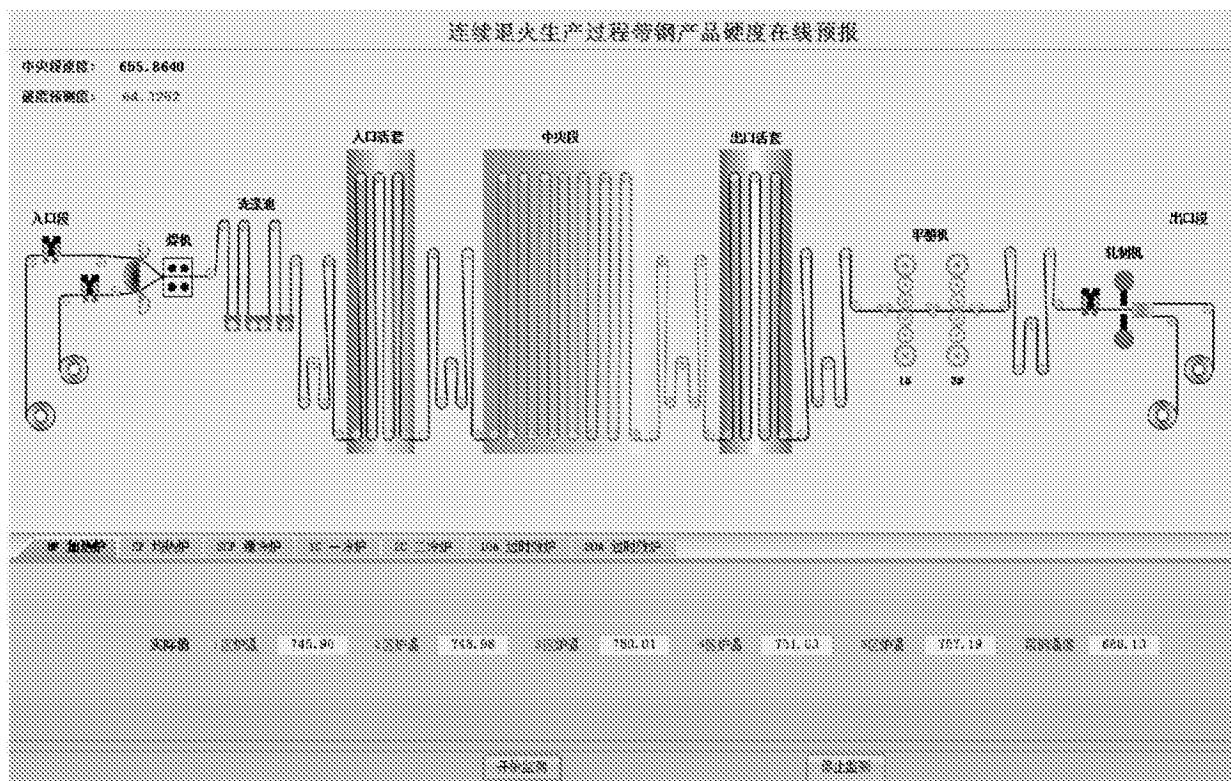


图5