

文章编号:1000-582x(2001)01-0095-04

基于粗糙集理论的连铸坯缺陷诊断预报系统

曹龙汉, 孙颖楷, 曹长修

(重庆大学自动化学院, 重庆 400044)

摘要:提出了基于粗糙集理论的连铸板坯缺陷类型的诊断及预报系统,以粗糙集理论为主要工具,揭示了板坯缺陷形成因素的冗余性,并对此约简后,利用径向基函数神经网络进行预报及诊断,初步结果表明该理论具有较好的实用性及有效性。

关键词:粗糙集;神经网络;聚类;缺陷诊断;预报

中图分类号:TP 391.76; TP 389.1

文献标识码:A

钢材连铸生产过程中,一些工艺参数经常偏离最佳状态,使得产品的质量有所下降,从而产生各类缺陷,对不同连铸坯质量的判断和预测,如裂纹、偏析、夹杂等,无疑是一线操作人员和管理人员都十分关心的问题。法国、日本、加拿大等国正在连铸生产线上使用的板坯缺陷诊断系统,根据结果提前预报,然后采取相应措施,对于减少因质量等级改判而引起的损失起到了不可估量的作用。但鉴于影响连铸质量各因素具有不确定性和非线性关系,许多成因仍在研究之中。而现有的板坯缺陷诊断系统都是基于从连铸生产机理出发所建立的预报模型,很难适应不同企业产品种类、原材料的变化,致使移植、维护变得非常困难。

本课题以连铸生产工艺为基础,利用人工神经网络、粗糙集理论等作为推理工具,建立板坯缺陷诊断及预报系统,以期帮助现场工作人员和管理人员及时调整工艺参数以确保连铸坯质量的稳定,从而降低因质量改判而带来的损失。

1 连铸坯主要缺陷及影响因素

根据英国钢铁研究学会的分类,连铸坯缺陷大致分为三大类31种,如图1所示。

由于板坯缺陷类型较多,考虑到表面缺陷造成的损失影响较大,故本课题首先从表面缺陷着手,选取表面纵裂、表面横裂、角部裂纹、表面夹渣4种类型加以

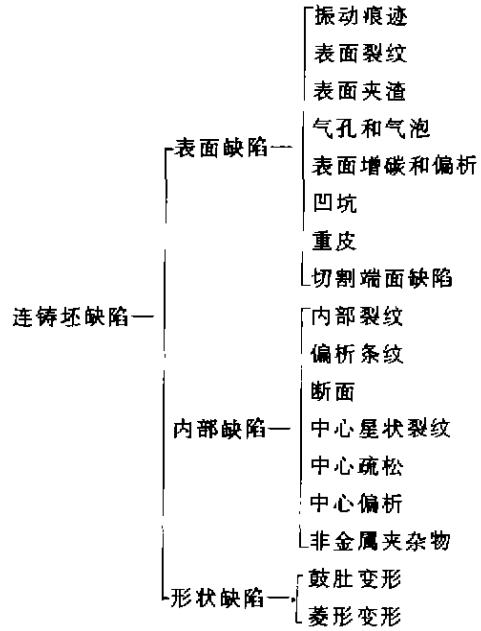


图1 连铸坯缺陷类型

研究。

表面纵裂是发生在板坯宽面中部与拉坯方向平行的裂纹,它是由于钢的高温应力超过钢的高温强度而在结晶器内产生的。该类缺陷造成板坯表面清理增大,收得率降低,严重时大量报废。它的产生与钢水成分中碳含量有很大关系,通常含碳量在0.1%~0.14%范围内,板坯容易产生表面纵裂;另外,钢水中S、

• 收稿日期:2000-06-21

基金项目:国家高等学校博士学科点专项基金(98061117)资助项目

作者简介:曹龙汉(1966-),男,四川省安岳县人,重庆大学博士生,现在重庆通信学院工作,教授。主要从事故障诊断、智能化通信电源方面的研究工作。

P、As 含量较高时,纵裂发生的几率增大;当高温浇注时,即使降低拉速仍会有纵裂纹产生,此外,保护渣的不均与流动会恶化板坯表面质量,二冷区设备状况也会对纵裂的产生造成一定影响。表面横裂、角部裂纹、表面夹渣等另外三种类型的形成因素也较多,在此不再赘述。

根据来自钢厂现场的调查报告,从中可以了解到连铸坯生产工艺中,影响不同钢种铸坯质量的工艺因

素很多,其中有些目前仍未弄清其原因。但主要的工艺因素有钢水成分、钢水温度、浇注速度、结晶器工作状态、结晶器冷却条件、水口安装与使用状况、保护渣性能以及二冷温度等。

由于缺陷成因的数量较多,数据量较大,表1只给出了某钢厂连铸坯表面缺陷影响因素的部分现场数据。

表1 连铸坯表面影响因素的现场数据

	C/%	Mn/%	P/%	S/%	Mn/S	F/%	液面波动 /mm	不均匀 指数/℃	结晶器 渣层/mm	...	类别
1	0.15	0.54	0.025	0.024	22	3.5	7.5	10	9		I
2	0.17	0.51	0.02	0.02	25	3.1	6.0	10	8		I
3	0.16	0.58	0.025	0.019	31	2.5	11	21	9		I
4	0.17	0.54	0.025	0.018	30	3.3	9.0	20	11	..	II
5	0.16	0.58	0.028	0.015	38	2.8	6.5	23	10		I
6	0.2	0.6	0.036	0.024	25	2.9	9.0	15	8		II
7	0.15	0.45	0.03	0.034	13	2.8	8.0	12	9		I
8	0.16	0.55	0.03	0.03	18	3.0	10.5	11	9		I
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴		∴

2 粗糙集理论的基本概念^[1]

粗糙集(Rough Set, RS)理论是由波兰华沙理工大学教授 Z. Pawlak 于 1982 年提出的,主要用于刻划不完整数据和不确定知识的表达、学习和归纳。它能有效地分析和处理不精确、不完整、不一致等不完备性,发现数据间隐藏的关系,从而提取有用信息,简化信息的处理。目前,RS 理论已在图象处理、模式识别、数据挖掘、专家系统等众多领域获得了成功的应用。

2.1 知识与不可分辨关系

粗糙集理论中,“知识”被认为是一种对抽象或现实的对象进行分类的能力,根据所讨论对象的特征差别将其分门别类的能力均可以看作是某种知识。在分类过程中,将相差不大的个体归为一类,它们之间的关系即为不可分辨关系(Indiscernibility relation, IND)。不可分辨关系是粗糙集理论的基石,揭示了论域中知识的颗粒结构,它是定义其他概念的基础,为此首先给出不可分辨关系的定义:

对决策系统 $S = (U, A)$, 每个属性子集 $B \subseteq A$, 称 $IND(B) = \{(x, y) \in U \times U: \forall a \in B, a(x) = a(y)\}$ 为 S 的不可分辨关系。

2.2 下近似集、上近似集和边界域

在 RS 理论中,通过一个不可分辨关系,对给定问

题的论域进行划分,得到不可分辨类 $[x]_{IND(B)}$, 然后确定其对某一概念的支持程度,分别以 3 个近似集合来表示,即下近似集、上近似集和边界域。具体定义如下:对每个概念 X , 包含于 X 中的最大可定义集称为 X 的下近似集(Low approximation),其定义为:

$$I_-(X) = \{x \in U: I(x) \subseteq X\}$$

包含 X 的最小可定义集称为 X 的上近似集(Upper approximation),定义为:

$$I_+(X) = \{x \in U: I(x) \cap X \neq \emptyset\}$$

包含于 X 的上近似而非 X 的下近似中的元素的集合,称为边界域,定义如下:

$$Bnd(X) = I_+(X) - I_-(X)$$

如果 $Bnd(X)$ 是空集,则称 X 关于 R 是清晰的;反之,如果 $Bnd(X)$ 不是空集,则称集合 X 为关于 R 的粗糙集。

2.3 决策表、约简和核

RS 理论使用决策表 KRS 来描述论域 U 中的对象。它是一张二维表格,每一行描述一个对象,每一列描述对象的一种属性。属性分为条件属性和决策属性,论域中的对象根据条件属性的不同,被划分到具有不同属性的决策类。对分类而言,并非所有条件属性都是必要的,有些是多余的,去掉这些属性不会影响原来的分类效果。约简定义为不含多余属性并保证分类

正确的最小条件属性集。一个决策表可能同时存在几个约简,这些约简的交集称为决策表的核,核中的属性是影响分类的重要属性。从另一个角度看,决策表中每个对象都蕴含着一条分类规则,化简决策表的过程也就是抽取分类规则的过程,利用化简后的决策表可以建立对象的专家系统。

3 连续属性值聚类量化^[2,3]

RS 理论使用量化属性值来优选条件属性,即进行最小约简,而影响连铸坯质量的因素,如钢水成分、浇注速度等都是连续属性,如何将其量化,是本课题首先需要解决的问题。对于连续属性值分类量化,可以归结为一个聚类问题,关于这方面的文献与方法较多,如多元统计分析、模糊聚类、神经网络等。

其中,SOFM 网络是一种无教师的竞争学习型前馈网络。基本原理为,当某类模式输入时,其输出层某一节点得到最大刺激而获胜,同时该获胜节点周围的一些节点因侧向作用也受到较大刺激,这时,与这些节点连接的权值矢量向输入模式的方向作相应的修正。当输入模式类别发生变化时,二维平面上的获胜节点也从原来的节点移到其他节点。这样,网络通过自组织方式用大量的训练样本数据来调整网络权值,最后使网络输出层特征图能反映样本数据的分布情况。因此,根据 SOFM 网络的输出状况,能得到整个数据区域的大体分布情况,即从样本数据中得到数据分布的大体本质特征。用 SOFM 网络量化连续属性值的步

骤如下:

① 初始化。对所有 n 个连接权值赋予一个较小的 $[0, 1]$ 区间的随机数;确定学习速率 $\eta(t)$ 的初值 $\eta(0)$ ($0 < \eta(0) < 1$);确定邻域的初值及总学习次数。

② 给定网络输入样本模式 X^k ;

③ 计算输入 X^k 与每一输出节点的距离:

$$d_i = \sum_{j=1}^m (x_j^k(t) - W_{ij}(t))^2, j = 1, 2, \dots, m$$

④ 选取距离最小的为获胜节点 j^* ;

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t)[x_j(t) - W_{ij}(t)],$$

$$\forall j \in NE_{j^*}(t), i = 1, 2, \dots, n$$

$$W_{ij}(t+1) = W_{ij}(t), \forall j \notin NE_{j^*}(t), i = 1, 2, \dots, n$$

邻域 $NE_{j^*}(t)$ 开始约为 $1/2$ 输出平面,中心点为 y_{j^*} ,以后逐渐压缩;

⑤ 对节点 j^* 及其邻域内的节点的连接权值进行更新;

⑥ 返回步骤 ②,重复执行,直到学习完毕。

本系统中,把钢水成份等多个连续属性量输入 SOFM 网络,根据分类结果,把不同的聚类分别用 1、2、3 等数字表示,例如碳含量(C%)小于 0.15 为 1,大于 0.15 为 2, ..., 就得到了连续属性的分类量化值。

4 连铸坯缺陷预报诊断系统^[4]

基于 Rough set 理论的连铸坯质量诊断及预报系统的实现框图如图 2 所示,具体步骤如下:

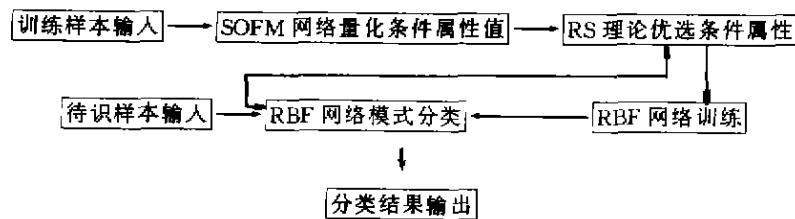


图 2 RS 智能网络系统框图

- 1) 用 SOFM 网络对连续属性值量化;
- 2) 用离散化后的条件属性(如钢水成份等)和决策属性(缺陷类型)建立决策表并使其相容;
- 3) 用 RS 理论对条件属性化简:
 - ① 删除决策表中相同的规则,即重复的实例;
 - ② 用 RS 理论的知识的简化方法,计算去掉一个条件属性后的上近似集,若相同,则删除表中该多余的条件属性;
 - ③ 用 RS 理论的范畴简化方法,若知识库 $K =$

(U, R) 中的基本范畴 $F = (X_1, X_2, \dots, X_n)$, 有 $\cap (F - \{X_i\}) = \cap F$, 则删除该条分类规则中的多余属性 X_i ;

④ 对每条规则进行简化,计算其决策规则的核值并建立核值表,求出最小约简;

⑤ 如果最后获得的核值表具有多个简化,可选取属性个数最少的简化进行下一步工作;

4) 利用优化的属性组合,采用径向基函数(Radial Basis Function, RBF)神经网络进行自组织模式分类

学习。

RBF神经网络是一种和BP网络类似的三层前向网络。输入层由信号源节点组成;第2层为隐层,变换函数是一种局部分布的对中心点径向对称衰减的非负非线性径向基函数;第3层为输出层。用RBF作为隐单元的“基”构成隐层空间,就可将输入矢量直接映射到隐层空间,当RBF的中心点确定后,这种映射关系就完全确定了。从隐层空间到输出空间的变换是线性的,即网络输出是隐单元的线性加权和。用RBF网络进行缺陷分类的步骤如下:

① 将简化后的条件属性作为RBF网络的输入,缺陷类型为输出,隐含层选择为高斯函数,即:

$$y_k = \exp(-(x - x_0)^2 / 2\delta^2)$$

② 用自组织学习方法选取RBF的中心点。通常采用C均值聚类算法,具体为:

a. 初始化聚类中心 C_j , ($j = 1, 2, \dots, L$, L 是 N 个样本中选取的 L 个样本);

b. 将输入样本按最邻近规则分组。如满足条件 $d_j = \min_{j=1}^L \|x_i - c_j\|$, 则将 x_i 分配给中心为 C_j 的输入样本聚类集合 θ_j , $j = 1, 2, \dots, L$;

c. 计算 θ_j 中样本的均值: $C_j = (\sum_{x_i \in \theta_j} x_i) / m_j$, m_j 为 θ_j 中的输入样本数,按上述步骤计算到聚类中心的分布不变为止;

d. 计算 δ : $\delta = \Delta m / \sqrt{2L}$

③ 输出层线性权值的调整,采用最小二乘法(RLS)进行学习。

5) 用训练好的RBF网络对经属性选择后的待识样本进行模式分类。

5 结束语

笔者将SOFM自组织神经网络、径向基函数RBF网络以及粗糙集理论应用于钢材板坯质量诊断及预报系统,以信息系统属性值表KRS为主要工具,去除了板坯缺陷成因的内在冗余性,得到一个具有属性个数最少的最小决策表。采用上述方法,用一批现场数据推导出连铸坯质量缺陷分类的最小决策表,并利用另一批现场数据样本进行验证,分类结果与实际样本的缺陷类型相吻合,表明粗糙集理论是一种较有效的工具。目前,本课题仍在继续进行中,它的应用为课题更深入的研究工作创造了有利条件。

参考文献:

- [1] PAWLAK Z. Rough Sets[J]. Communication of the ACM, 1995, 38(11): 89-95.
- [2] 李永敏. 基于粗糙集理论的数据挖掘模型[J]. 清华大学学报(自然科学版), 1999, 39(1): 110-113
- [3] 张萍, 王桂增, 周东华. 动态系统的故障诊断方法[J]. 控制理论与应用, 2000, 17(2): 153-158.
- [4] 王永骥. 神经元网络控制[M]. 北京: 机械工业出版社, 1998.

Slab Defect Diagnosis and Predicting System Based on Rough Set Theory

CAO Long-han, SUN Ying-kai, CAO Chang-xiu

(College of Automation, Chongqing University, Chongqing 400044, China)

Abstract: This paper presents the continuous casting slab defect diagnosis and predicting system based on rough set theory. The knowledge represent system is used as a tool to unveil the redundancy of slab's defect. Radial basis network is used to predict and diagnose the type of defect after reduced the redundancy. The results show that the rough set theory is practical and effective.

Key words: rough set; neural network; clustering; defect diagnosis; predict

(责任编辑 吕赛英)