

doi:10.11835/j.issn.1000-582X.2014.03.010

基于粗糙集的转炉炼钢知识发现模型

胡 燕, 郑 忠

(重庆大学 材料科学与工程学院, 重庆 400044)

摘 要:针对转炉炼钢知识发现的特点,采用粗糙集理论进行分析,应用数据清洗、标准化及离散等方式对转炉炼钢生产数据进行预处理,以炼钢生产的主要影响因素作为知识发现的条件属性,以转炉冶炼终点控制目标作为知识发现的决策属性,建立了基于粗糙集方法的转炉炼钢知识发现模型,实现转炉炼钢生产知识的自动发现、获取和规则提取。以转炉冶炼终点钢水温度的变化规律做为知识发现的决策属性,采用 210 t 转炉炼钢实际生产数据进行模型的应用测试,结果表明提取出的铁水硅含量、铁矿石质量、氧气消耗量等影响因素对转炉冶炼钢水终点温度存在重要影响,且模型提取出的转炉炼钢终点钢水温度知识规则与现行转炉炼钢现场的变化规律一致,证明基于粗糙集方法的转炉炼钢知识发现模型的有效性。

关键词:知识发现模型;粗糙集;转炉炼钢

中图分类号: TF345.3

文献标志码: A

文章编号: 1000-582X(2014)03-058-06

Knowledge discovery model of basic oxygen furnace steelmaking based on Rough Set Theory

HU Yan, ZHENG Zhong

(College of Materials Science and Engineering, Chongqing University, Chongqing 40044, China)

Abstract: The characteristics of the knowledge discovery for basic oxygen furnace (BOF) steelmaking are analyzed by using the Rough Set Theory. The production data of BOF steelmaking are preprocessed by using the methods of data withdrawal, standardization, discretization and so on. The main influencing factors of steelmaking production are set as the knowledge discovery property. The endpoint control objectives of BOF steelmaking are used as the decision attribute of knowledge discovery. Then the knowledge discovery model of BOF steelmaking based on rough set theory is established, which makes the automation of the production knowledge discovery, access and rule extraction come true. The model is tested by using the production data of 210 t BOF, and takes the temperature variation of smelting endpoint as the decision attribute. The results show that the influencing factors, such as silicon content, iron ore weight, oxygen consumption and so on, are of very importance to the endpoint temperature of molten steel. Besides, the rules of molten steel temperature extracted by the model vary with current converter steelmaking process, which proves the validity of the model.

Key words: knowledge discovery model; rough sets methodology; basic oxygen furnace steelmaking

收稿日期:2013-10-15

基金项目:国家自然科学基金资助项目(51274264)

作者简介:胡燕(1981-),女,重庆大学博士研究生,主要研究方向为过程工程控制技术。

郑忠(联系人),女,重庆大学教授,博士生导师,(E-mail)zhengzhong@cqu.edu.cn。

转炉炼钢是一个复杂的多元多相高温物理化学过程,其控制核心是对冶炼终点的钢水温度和碳含量进行准确的预测和判断^[1-2]。目前,国内绝大多数钢铁厂的转炉炼钢终点控制还是依据现场操作人员的生产操作经验进行,由于现场技术人员的素质参差不齐,实际生产工况条件又很复杂,使得仅靠人工经验得到的转炉冶炼知识进行终点控制时,其稳定性和准确性不能满足炼钢生产的要求。

随着转炉炼钢过程控制技术的发展,转炉炼钢计算机服务器中采集了大量的生产数据信息,但是,炼钢生产过程中采集的数据信息普遍存在多变量、非线性、高噪音的特征,因此,如何从大量的数据中找出有效实用的知识,并用于指导炼钢终点控制技术具有重要的意义^[3]。粗糙集方法基于不可分辨性的理论思想,可以描绘知识表达中不同属性重要性,实现从数据中推理逻辑规则作为知识系统的方法,非常适用于转炉炼钢这样的过程复杂、很难用数学模型描述的大型工业生产问题。同时,针对转炉炼钢生产过程控制中积累的庞大数据资源,利用粗糙集方法可以对过程数据进行有效的数据特征分析,减少数据噪声和无关数据对转炉终点控制目标的影响,保留关键变量,提取数据库中有用的知识,进而优化转炉炼钢终点控制方法,提高转炉终点控制的准确性。

1 粗糙集理论方法及知识发现流程

1.1 粗糙集方法原理

粗糙集方法是基于不可分辨性思想和知识简化,对不完整数据进行分析、推理、学习、发现,进而从数据中推理逻辑规律的智能知识发现方法^[4-7]。粗糙集方法认为知识都是有粒度的,正是因为知识的颗粒性造成使用已有知识不能精确地表示某些概念,从而形成知识的不确定性,因此,粗糙集理论通过不可区分关系和等价类划分所确定问题的近似域(上近似和下近似),能有效分析和处理不精确、不完整和不一致等各种不完备数据,通过知识约简,推导出概念的分类规则,从而发现隐含知识,揭示数据中潜在的规律^[8-12]。

假设给定知识库 $K = (U, R)$, 其中 U 表示论域,而 R 为等价关系,对于每个子集 $X \in U$ 和一个等价关系 $R \in \text{ind}(K)$, 可以根据 R 的基本集合描述来划分集合 X 。为了衡量 R 对集合 X 描述的精确性,考虑两个子集

$$R_-(X) = U\{Y \in U/R; Y \subseteq X\}, \quad (1)$$

$$R_+(X) = U\{Y \in U/R; Y \cap X \neq \emptyset\}, \quad (2)$$

$$b_{n_R}(X) = R_+(X) - R_-(X), \quad (3)$$

其中: $R_-(X)$ 为 X 的 R 下近似集; $R_+(X)$ 为 X 的 R 上近似集; $b_{n_R}(X)$ 为 X 的 R 边界。

$\text{pos}_R(X) = R_-(X)$ 称为 X 的正域,是知识 R 能完全确定归入集合 X 的对象集合; $\text{neg}_R(X) = U - R_+(X)$ 称为 X 的 R 负域,表示知识 R 确定不属于集合 X 的元素集合; 边界域 $b_{n_R}(X) = R_+(X) - R_-(X)$ 是论域中不确定域,根据知识 R 属于边界域的元素对象不能确定地划分集合 X 或是 X 的补集。

1.2 基于粗糙集方法的知识发现流程

基于粗糙集方法的知识发现是将用于知识发现的数据信息转换为决策表,通过对决策表中的条件属性约简形成属性约简表,最后由属性约简表发现数据信息中潜在的知识 and 规律的过程(如图 1 所示)。

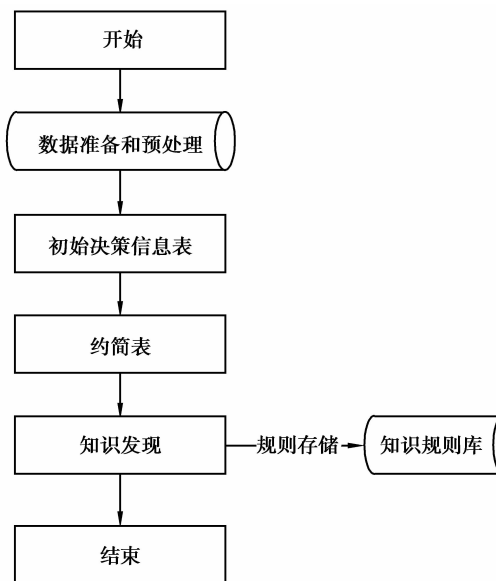


图 1 基于粗糙集理论的知识发现流程

1.2.1 数据准备和预处理

数据准备是根据用户需求从原始数据库中抽取确定的目标数据。数据预处理一般包括消除噪音、推导计算缺值、消除重复记录、完成数据类型转换等。

1.2.2 初始决策信息表

决策信息表是由条件属性和决策属性构成一个多维表征用户知识发现需求的初始特征表。在初始决策信息表中,决策属性就是知识发现的目标,条件属性是有可能对决策属性产生影响的因素。

2.2.2 转炉炼钢知识发现的决策表

确定转炉炼钢知识发现的条件属性和决策属性后(如表 1 所示),采用数据删除法(解决不完备数据和冗余数据)、多数优先原则(解决不一致数据)及数据归一化处理(消除数据量纲影响)等数据清洗方法对条件属性和决策属性数据集进行清洗后,形成一个多维表征转炉炼钢生产工艺特点的初始特征表,如表 2 所示。

表 2 部分条件属性与决策属性构成的初始特征样表

铁水硅含量/ %	铁水温度/ ℃	石灰消耗量/ kg	氧气消耗量/ Nm ³	吹炼时间/ s	钢水碳含量/ %	钢水温度/ ℃
0.49	0.81	0.47	0.54	0.56	0.49	0.66
0.36	0.35	0.30	0.61	0.65	0.34	0.72
0.44	0.52	0.26	0.65	0.69	0.38	0.57
0.53	0.90	0.30	0.47	0.54	1.00	0.52
...

初始特征表反应铁水 Si 含量、铁水温度、石灰消耗、氧气消耗等条件属性与决策属性钢水碳含量和钢水温度的变化关系。虽然初始特征表中可以通过统计方法得出相应的条件属性参数特征与决策属性的关系,但初始特征表的信息是具以数值形式表示,很难从中得出其代表具体含义,因此,需要对初始特征表进行概念提升,将特征表中特征的数值性描述转化为人们容易理解的概念性描述,从而形成用于约简的决策表。

提取数据信息的数值型概念层次需具备两个特性:完整性和均匀性,因此,为了满足以上特性,笔者采用等频率的概念提升方法,通过将属性值域划分成若干区间,使用基于概念树的属性区间自动分类算法自动划分属性的值域区间,保证落入每一区间的数值分布频率相等。同时,针对转炉炼钢的数据特点,将属性的概念区间划分为 7 个:L(很低)、LL(低)、L(偏低)、C(正常)、H(偏高)、HH(高)、HHH(很高),通过概念提升,将初始特征样表转换为决策信息表(如表 3 所示),使得决策表中各组成属性都有特定含义的概念,便于知识规律的发现。

表 3 决策信息样表

铁水硅含量/ %	铁水温度/ ℃	石灰消耗量/ kg	氧气消耗量/ Nm ³	吹炼时间/ s	钢水碳含量/ %	钢水温度/ ℃
C	HHH	C	C	C	C	HH
LL	LL	LL	H	H	LL	HHH
L	C	LL	HH	H	LL	C
C	HHH	LL	C	C	HHH	C
...

2.2.3 决策信息表的属性约简

决策信息表中蕴涵了很多有用的决策信息和知识规律,但也存在很多的噪音信息(冗余属性)进而影响系统决策的准确性。信息系统理论提供了一种有效去除冗余属性的约简方法,即以属性重要性程度对属性进行约简,其原理是将条件属性对决策属性的重要性程度进行量化度量,并按照从大到小进行排序,将重要性程度较大的属性集作为影响决策变量的主要影响因素,从而完成属性约简的过程。

笔者采用基于属性重要性的约简方法对冗余属性进行约简:以核作为计算约简的出发点,以属性的重要性作为启发规则,在求出决策表核的基础上求得属性的重要性,并对核以外的属性按重要性由大到小的顺序排列,然后,按照这一顺序将其添加到核心属性集中,直到该集合是一个约简为止。

2.2.4 基于关联规则的知识发现和规则提取

关联规则提取方法是在约简表的基础上归纳出条件属性和决策属性之间的关联关系。关联规则的知识发现和规则提取中,规则的前提对应一个特征集描述 X,规则的结论对应一个概念描述 Y,则定量关联规则可表示为 X→Y [S][C],其中 S,C 分别代表规则的支持度和信任度。

$$S = \frac{\text{card}([X] \cap [Y])}{\text{card}(U)}, \quad (4)$$

$$C = \frac{\text{card}([X] \cap [Y])}{\text{card}([X])}. \quad (5)$$

关联规则提取的目标是在约简表中找出满足用户最小支持度和最小信任度的规则。由于通过计算每个特征集的描述与概念描述间的支持度和信任度可以发现出所有的关联规则,但是这种方法存在计算效率低、提取规则存在大量冗余、规则解释能力低的缺点,因此,为了提高计算效率,避免冗余规则的产生,笔者采用高效的关联规则提取算法发现转炉炼钢数据中的生产知识规则^[14-15],如表 4 所示。

表 4 基于粗糙集方法提取的转炉炼钢知识规则

序号	前提数	前提描述	结论描述	支持度	信任度
1	3	12&33&61#	86#	0.3	0.5
2	3	10&30&70#	86#	0.3	0.3
3	2	4&70#	85#	0.3	0.6
4	3	37&69&64#	84#	0.3	0.3
...

说明:前提号与描述对应为:4(铁水碳含量正常)、10(铁水硅含量偏低)、12(铁水硅含量偏高)、33(铁水温度偏高)、61(铁矿石质量低)、30(铁水温度低)、70(氧气消耗量正常)、37(铁水质量正常)、69(氧气消耗量偏低)、64(铁矿石质量偏高)、84(钢水终点温度偏低)、85(钢水终点温度正常)、86(钢水终点温度偏高)。

3 模型应用

笔者以某 210 t 转炉炼钢实际生产数据作为炼钢知识发现模型应用的基础数据,以转炉炼钢冶炼终点的钢水温度变化为目标,采用粗糙集知识发现模型实现对转炉炼钢知识规则的提取(冶炼终点钢水温度和碳含量的知识规则提取方法相同),如表 5 所示。

表 5 转炉炼钢知识规则及其解释

规则号	规则	定量结论	S	C	操作规程 工艺解释
1	IF 实际铁量硅含量偏高 AND 铁水温度偏高 AND 铁矿石加入量低 THEN 钢水终点温度偏高	[1 700,1 751]	0.3	0.5	定性一致
2	IF 实际铁水量正常 AND 氧气消耗量偏少 AND 铁矿石加入量偏高 THEN 钢水终点温度偏低	[0,1 665)	0.3	0.3	定性一致
3	IF 实际铁水碳含量正常 AND 氧气消耗量正常 THEN 钢水终点温度正常	[1 665,1 700)	0.3	0.6	定性一致
4	IF 石灰消耗正常 AND 生白云石消耗高 THEN 轻烧白云石消耗正常 THEN 钢水终点温度低	[0,1 665)	0.4	0.4	定性一致
...

由表 5 可以看出,提取出来的转炉炼钢生产知识与现行转炉炼钢现场的终点钢水温度变化规律一致。其中,提取出的铁水碳含量、铁水硅含量、铁矿石重量、氧气消耗量对转炉冶炼钢水终点温度存在的影响规律完全符合转炉冶炼的生产操作工艺,这也说明此知识发现算法中基于极大信息量理论的数据约简没有将一些明显的重要属性进行删除,同时证明该知识约简方法对影响转炉冶炼钢水终点温度因素约简的有效性。

4 结论

针对复杂转炉炼钢过程潜在生产知识规则发现的难题,引入粗糙集理论方法,建立基于粗糙集的转炉炼钢知识发现模型,并以具体炼钢转炉生产过程数据进行了方法的验证及应用测试,表明了基于粗糙集方法的转炉炼钢知识发现模型的有效性,同时,该研究方法也为转炉炼钢知识发现研究和生产控制提供了新方法和新手段。

参考文献:

- [1] 何平,刘浏. 赵进宣. 转炉炼钢综合智能型静态控制模型改进与应用[J]. 钢铁研究, 2012, 40(6): 22-26.
HE Ping, LIU Liu, ZHAO Jinxuan. Improvement and application of comprehensive and intelligent models for static control of steelmaking in converter [J]. Research on Iron & Steel, 2012, 40(6): 22-26.
- [2] 姬厚华. 120 t 转炉炼钢模型控制系统的研究与应用[J]. 冶金动力, 2008(4): 77-79.
JI Houhua. Study and of 120t application of control converter steelmaking system model [J]. Metallurgical Power, 2008(4): 77-79.
- [3] 菅希顺,刘瑞霞. 数据挖掘技术及其在钢铁领域应用概述[J]. 天津冶金, 2006(1): 39-42.
Jian Xishun Liu Ruixia. Summerization of the data excavation technology and its application in iron and steel industry [J]. Tianjin Metallurgy, 2006(1): 39-42.
- [4] Liu W J, Guo Q. An algorithm to select the optimal program based on rough sets[J]. Journal of

- Convergence Information Technology, 2012, 7(23): 442-449.
- [5] Sengupta S, Das A K. Dimension reduction using clustering algorithm and rough set theory [J]. Lecture Notes in Computer Science, 2012, 7677: 705-712.
- [6] Herawan T, Yanto I T R, Deris M M I. Rough set approach for categorical data clustering [J]. International Journal of Database Theory and Application, 2009, 66: 33-52.
- [7] 王希雷, 马永军, 苏静. 基于 Rough 集的数据挖掘中知识变化的研究 [J]. 华中科技大学学报: 自然科学版, 2012, 40(Sup1): 320-323.
WANG Xilei, MA Yongjun, SU Jing. Research on knowledge change of data mining based on rough sets [J]. Journal of HuaZhong University: Science and Technology: Natural Science Edition, 2012, 40(Sup1): 320-323.
- [8] 郑芸芸, 王萍, 游强华. 一种基于粗糙集的数据挖掘模型 [J]. 软件导刊, 2012, 11(11): 102-104.
ZHENG Yunyun, WANG Pin, YOU Qianghua. A data mining based on rough set theory [J]. Software Guide, 2012, 11(11): 102-104.
- [9] 陈秀明, 王璐. 基于粗糙集方法的信息系统中属性划分方法研究 [J]. 中北大学学报: 自然科学版, 2012, 33(6): 695-698.
CHEN Xiuming, WANG Lu. Research on classification of attribute in information system based on rough set theory [J]. Journal of North University of China: Natural Science Edition, 2012, 33(6): 695-698.
- [10] Greco S, Matarazzo B, Stowinski R. Dominance-based rough set approach on pairwise comparison tables to decision involving multiple decision makers [J]. Lecture Notes in Computer Science, 2011, 6954: 126-135.
- [11] Chen C, Wang L. Integrating rough set clustering and grey model to analyse dynamic customer requirements [J]. Journal of Engineering Manufacture, 2008, 222(2): 319-332.
- [12] SUN L X, Li L L. Optimization with rough set for system call sequence analysis in intrusion detection [J]. International Journal of Digital Content Technology and its Applications, 2012, 6(23): 31-38.
- [13] 刘胜军, 杨学兵, 蔡庆生. 关系数据库中概念层次自动提取算法研究 [J]. 计算机应用研究, 1999(12): 15-17.
LIU Shengjun, YANG Xuebin, CAI Qingsheng. The automatically extraction algorithm of concept hierarchy in relational database [J]. Computer Application Research, 1999(12): 15-17.
- [14] 崔旭, 刘小丽. 基于粗糙集的改进 Apriori 算法研究 [J]. 计算机仿真, 2013, 30(1): 229-332.
CUI Xu, LIU Xiaoli. Improved apriori algorithm based on rough set [J]. Computer Simulation, 2013, 30(1): 229-332.
- [15] 陈波, 邵勇, 王成华, 等. 高效的关联规则挖掘算法研究 [J]. 计算机工程与设计, 2008, 29(24): 6240-6242.
CHEN Bo, SHAO Yong, WANG Chenghua, et al. Research on high efficient algorithm for mining association rules [J]. Computer Engineering and Design, 2008, 29(24): 6240-6242.

(编辑 张 苹)