

# CBR 及 KDD 技术在中厚板轧制 负荷分配建模中的应用

马当先<sup>1)</sup> 穆志纯<sup>2)</sup> 王绍波<sup>2)</sup> 王长松<sup>1)</sup>

1)北京科技大学机械工程学院,北京,100083 2)北京科技大学信息工程学院,北京 100083

**摘要** 针对中厚板生产产品规格变化大、轧制条件复杂等特点,提出一种基于案例推理的负荷分配模型,并应用数据库中的知识发现技术进行案例修正,使模型更符合实际轧制情况。

**关键词** 负荷分配模型;基于案例的推理;数据库中的知识发现

**分类号** TP 181; TG 335.5'1

中厚钢板是工业生产中不可缺少的重要材料,被广泛用于国防、交通运输、能源和建筑等各种重要的国民经济生产部门<sup>[1]</sup>。在中厚板轧制中,板形是一个非常重要的质量指标,它是保证生产正常运行的重要条件。在国内的中厚板轧制实践中,负荷分配是控制板形的主要手段之一。但中厚板生产产品规格变化很大,轧制条件非常复杂,能适应各种实际生产工况的负荷分配模型的建立相当困难。因此,建立合理的负荷分配模型一直是中厚板轧制技术的研究热点。本文将 CBR (Case-Based Reasoning, 基于案例的推理) 与 KDD (Knowledge Discovery in Database, 数据库中的知识发现) 技术相结合,提出了一种新的建立负荷分配模型的思路和方法,并对模型结构及具体处理过程进行了介绍。利用现场数据进行的对比实验计算结果表明,这种方法是有效的,较目前流行的负荷分配模型的计算结果有了改进。

## 1 建模方案的选择

负荷分配模型对板形的影响主要体现在对轧制的最后几个道次的负荷分配上。因此,不同负荷分配模型的主要差异在于对轧制最后几个道次的负荷分配所采用的原则和方法。对于这几个道次的负荷分配,一种方法是按成品典型宽度和典型厚度组成的二维表格来存储其最后几道次的轧制力及出口厚度的设定值,得到一簇 PH

曲线(称为 PH 图模型),使用时再根据待轧钢坯的成品尺寸进行插值计算获得其最后几道次的应用 PH 曲线<sup>[2]</sup>。这种方法思想简单,但建立包含各种钢种、规格的二维表格的工作量巨大,且需要有足够的现场条件,开发周期长;而利用插值计算对典型规格产品负荷分配的修正,很难保证对各种规格的产品均具有满意的控制效果。另一方面,若利用操作工人的丰富操作经验,采用传统的基于规则推理的方法建立负荷分配模型也是一个较好的思路,但它存在着实现上的困难。首先,将操作工人的丰富经验全面地转化为规则绝非易事;其次,基于规则推理的系统非常脆弱,必须完全满足规则条件才能解决问题,如果问题处于系统已有规则的范围之外,系统对这样问题的求解将无能为力。

CBR 技术和 KDD 技术的有机结合能够克服传统的基于模型推理与基于规则推理的不足,如知识获取出现瓶颈,对求解问题的历史缺乏记忆,需要靠专家事先经过大量的实验和推导才能形成模型和规则,处理问题存在知识边界等。本文工作的思路是,利用 CBR 的思想和技术建立轧制规程案例库,利用 KDD 技术从轧制运行数据库中发现案例修改规则以适应不同轧制条件和不同轧制规格,由此建立一种基于经验和知识模型的中厚板轧制负荷分配方法。利用该方法建立模型,计算轧件最后几道次的负荷分配曲线作为道次设定的依据,使负荷分配模型适应性更好,更符合应用现场情况。

基于案例推理是一种类比推理方法,其一般

收稿日期:2003-10-13 修回日期:2004-03-10

作者简介:马当先(1959—),男,教授

推理过程为:问题描述→案例检索→案例修改→案例保存.首先从用户那里得到对问题(案例)的描述,根据案例描述,在案例库中进行检索,找到最佳匹配的案例.再根据此案例和问题描述的差异,对此案例进行修正,使之适应当前问题的描述.最后,将问题的解提供给用户,并将此新案例存入案例库<sup>[3]</sup>.其中,案例保存既是学习的过程,也是知识获取的过程<sup>[4]</sup>.在 CBR 系统中,最困难的一步是案例修改,它同具体问题的领域知识有关.然而,不少 CBR 应用系统没有进行案例修改的有效机制,将修改的任务完全推给了用户.在解决本文研究的特定问题中,笔者利用 KDD 技术从现场采集的大量轧制数据(轧制运行数据库)中挖掘潜在的领域知识,发现案例修改加权关联规则,在 CBR 原始案例的基础上,根据这些修改规则进行修正,以适应不同的轧制情况.

## 2 模型结构与实现方法

本文提出的基于 CBR 和 KDD 的负荷分配模型结构如图 1 所示.

(1)新案例的输入与表示.案例的表示是指用一种适当的形式描述案例的特征,将案例在计算机中表示出来,它是基于案例推理的前提和基础.本文采用面向对象技术进行案例的表示,将案例的属性数据以及案例检索、修改、保存等方法封装在案例类中,系统在运行中动态地构造案例对象.

根据本文研究问题的目的和特点,选取轧制过程最后几道次(例如最后 6 个道次)的轧制力以及出口厚度作为案例的决策属性.实际轧制中影响负荷分配的因素很多,综合考虑各种因素影响的大小及检索模型的规模并结合工人的轧制经

验,选择钢种、平均辊径、坯料长/宽/厚、成品宽/厚、开轧温度、终轧温度 9 个属性作为案例的条件属性. CBR 模型的输入是案例的条件属性,输出是案例的决策属性.

(2)案例检索.案例检索的实质就是从案例库中寻找与当前案例最为相似案例.常用的案例检索方法主要有三种:最近邻法,归纳索引法和知识导引法<sup>[5]</sup>.本文采用最近邻法的思想计算案例的相似程度,即:

$$S(C_i, C_j) = \sum_{k=1}^n \omega_k \frac{a_{ik}}{a_{jk}} \quad (1)$$

其中,  $\omega_k$  为案例第  $k$  个条件属性的权值,其值根据经验按对负荷分配影响的大小设定;  $a_{ik}, a_{jk}$  分别为案例  $C_i$  与案例  $C_j$  的第  $k$  个条件属性值.特别地,钢种属性是一种类别属性,若两个案例的钢种属于同一类,则该属性的相似度  $\frac{a_{ik}}{a_{jk}}$  取 1,否则取 0.

最近邻法简单实用,但对于每个输入的当前案例需要和案例库中的每个案例求相似度,对于大型案例库检索效率低.因此在实际应用中,利用知识对最近邻法进行了简化.由于成品厚度对最后几个道次负荷分配的影响最大(成品厚度越薄越易出现板形问题),因此在案例检索的时候,让当前案例只与案例库中成品厚度属性值相等的案例进行相似度的比较,这样大大减少了需要计算相似度的案例的数量,提高了案例检索的速度,满足了应用背景的实时性要求.

(3)案例修改. CBR 系统检索到的案例往往只是与当前问题类似,但还有一定差距,需进行修改.案例修改是 CBR 应用中的一个难点,对领域知识依赖性很强.

本研究的重点在于求解轧件最后几道次合理的负荷分配曲线.根据经验和实际轧制数据分析,最后几个道次的轧制压力下降曲线近似于一条直线.因此可将案例修改规则简化为由钢坯的几个重要属性参数值推出合理的终轧压力及最后几个道次中每道次轧制压力的下降量,即简化为  $(SteelKind=a) \wedge (Shplt=b) \wedge (Swplt=c) \wedge (EndT=d) \Rightarrow (EndP=e) \wedge (EndPSlope=f)$ . 其中, SteelKind, Shplt, Swplt, EndT 分别表示钢种、成品厚度、成品宽度、终轧温度,是规则的条件属性; EndP, EndPSlope 分别表示终轧压力及压力下降量,是规则的结果属性.这种规则与 KDD 中关联规则的情形比较类似,因此可以通过挖掘轧制运行数据库中的关联规则获取案例修改的规则.

设  $I = i_1, i_2, \dots, i_m$  是  $m$  个不同交易项目的集合,给定一个交易数据库  $D$ ,其中每个交易  $T \subseteq I$  都与

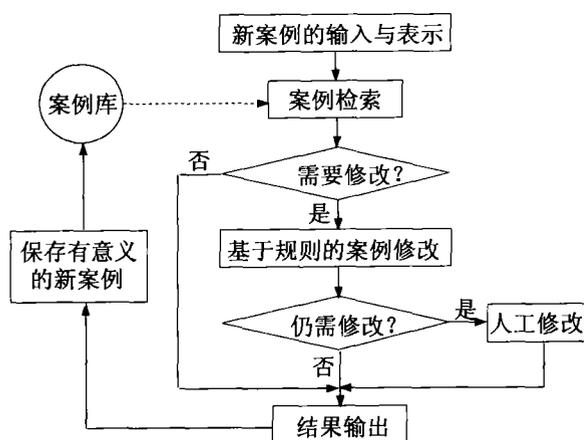


图 1 基于 CBR 的负荷分配模型结构

Fig.1 Structure of CBR-based schedule model

一个唯一的标识符 TID 相关联. 如果对于  $I$  中的一个子集  $X$ , 有  $X \subseteq T$ , 称交易  $T$  支持  $X$ . 一条关联规则就是一个形如  $X \Rightarrow Y$  的蕴含式, 其中  $X \subseteq I, Y \subseteq I$ , 且  $X \cap Y = \emptyset$ . 其意义在于一个交易中某些项目的出现, 可推导出另一些项目在同一交易中也出现. 若  $D$  中有  $s\%$  的交易同时支持项目集  $X$  和  $Y$ ,  $D$  中支持项目集  $X$  的交易中有  $c\%$  的交易同时也支持项目集  $Y$ , 称规则  $X \Rightarrow Y$  在  $D$  中具有  $s\%$  的支持度,  $c\%$  的置信度. 关联规则的挖掘问题就是在给定的交易数据库  $D$  中发现满足用户给定的最小支持度和最小置信度的关联规则的过程.

按处理对象的不同, 关联规则可分为布尔型关联规则和多值关联规则<sup>[6]</sup>. 布尔型关联规则挖掘问题的主要应用对象是由布尔型属性组成的交易数据库, 是在给定的交易数据库中发现满足用户给定的最小支持和最小置信度的布尔型关联规则的过程. 多值关联规则挖掘问题的主要应用对象是由多值属性组成的记录集, 是在给定的记录集中产生满足用户给定的最小支持度和最小置信度的多值关联规则的过程. 多值属性主要有数量属性(如年龄、工资等)和类别属性(如性别、学历等).

本文所要挖掘的案例修改规则中的属性值既有数量属性(成品尺寸、终轧温度、终轧压力等), 又有类别属性(钢种), 因此可以按照 QARP 的处理方法进行规则挖掘. 目前已有的关联规则挖掘的许多算法都是针对 BARP 提出的, 主要研究对象是项目集类型的数据集(如超市中每个顾客购买的商品形成一个项目集), 通过对数据库的多次扫描挖掘出所隐含的项目及其数量预先未知的关联规则. QARP 比较复杂, 一般先将它转化为 BARP, 再使用 BARP 的挖掘算法. 由于笔者在应用中利用领域知识事先确定了与挖掘任务有关的因素, 明确了要挖掘的规则的形式, 因而可以避免挖掘的盲目性, 采用一种更为有效的特殊挖掘算法. 这种算法首先进行布尔化转换, 将数据的各个属性按一定标准划分为一个个单位区间, 将每个区间按顺序映射为一个非负整数, 并将每一条记录的属性值映射为非负整数. 属性区间可以按数据的分布特征划分, 也可以按  $n$  等分的标准划分, 或按等频的标准进行划分. 从空间的角度来看, 全部条件属性构成了一个多维数据空间. 在属性区间划分后, 各维度的单位区间在多维空间中交叉构成了一个个被称为数据包 (data cube)<sup>[7]</sup> 的小单位空间, 而每一条数据记录可

以看作数据包中的一个点. 通过数据包, 重新定义支持度和置信度, 更便于关联规则的挖掘.

上述算法中, 布尔化转变是非常关键的一步. 区间划分太窄可能使一些规则因其支持度达不到最小支持度而丢失, 产生“最小支持度”问题; 区间划分太宽可能使一些规则因其置信度达不到最小置信度而丢失, 产生“最小置信度”问题. 根据工人实际轧制经验, 终轧厚度对终轧压力的取值影响较大, 因此对于终轧厚度的属性划分区间不能过大. 特别是成品厚度很薄时, 对板形要求更高, 划分应更细. 为体现薄板数据的重要性, 同时避免布尔化过程中的“最小支持度”问题, 采用加权方法, 按成品厚度赋予权值, 对包含较小成品厚度的记录赋予较大的权值, 增加其加权支持度. 以规则  $(SteelKind=0) \wedge (Shplt=1) \wedge (Swplt=0) \wedge (EndT=4) \Rightarrow (EndP=5) \wedge (EndPSlope=3)$  为例, 设  $Shplt=1$  对应权值 0.9, 通过数据包, 其加权支持度与加权置信度分别定义如下.

① 加权支持度: 数据包 (0,1,0,4) 中  $EndP=5$  同时  $EndPSlope=3$  的记录数同 0.9 的乘积与加权记录总数的比值.

② 加权置信度: 数据包 (0,1,0,4) 中  $EndP=5$  同时  $EndPSlope=3$  的记录数与该数据包中包含的记录总数的比值.

其中, 加权记录总数定义为成品厚度落在其各布尔化单位区间上的记录数与该区间对应权值的乘积的总和.

对每个数据包进行加权计算, 可以很方便地从中发现案例修改规则. 由于挖掘的修改规则只是一种简化形式, 忽略了一些其他因素, 因此在完成了基于规则的案例修改之后, 得到的软件负荷分配曲线有可能仍需要进一步修改. 这时, 可提供人机接口界面, 由操作工直接进行调整.

(4) 案例保存. 完成案例修改后, 得到当前案例的所有属性值, 即一个完整的新案例. 若用户觉得该案例有保存价值, 则将其存入案例库, 以备后用; 如果用户觉得以后不需要或者该案例可由现在库中已有的案例很方便地再生出来, 则可不进行存储. 通过保存有意义的新案例, 使案例库得到不断的补充完善, 体现了基于案例推理优于其他传统推理方式的学习能力.

### 3 对比实验结果

应用本文提出的基于 CBR 和 KDD 的负荷分配模型进行了在线对比实验计算, 得到了比较合

理的规程。以 Q235A 钢为例, 平均辊径 950 mm, 坯料厚×宽×长=150 mm×1 200 mm×3 300 mm, 成品厚×宽=10 mm×2 300 mm, 开轧温度 1 035℃, 终轧温度为 850℃, 分别用基于 CBR 和 KDD 的负荷分配模型和 PH 图模型进行计算, 将得到的规程与在某厂现场实际采集统计得到的规程相比较, 最后 6 道次的辊缝值如表 1 所示。

表 1 规程比较 (辊缝值)

Table 1 Comparison of the schedules (Roll gap) mm

| 道次 | 利用 PH 图模型生成的规程 | 基于 CBR, KDD 生成的规程 | 实际采集得到的规程 |
|----|----------------|-------------------|-----------|
| 1  | 6.31           | 6.61              | 6.54      |
| 2  | 7.93           | 7.39              | 7.25      |
| 3  | 9.88           | 8.79              | 8.79      |
| 4  | 11.93          | 11.78             | 11.59     |
| 5  | 15.15          | 14.26             | 14.03     |
| 6  | 19.63          | 19.54             | 19.18     |

从表 1 中的辊缝数据中可以看出: 利用 PH 图模型生成的规程中, 最后几个道次, 尤其是最终道次, 压下量太大, 无法得到符合要求的板形; 由基于 CBR 和 KDD 的轧制规程计算模型生成的规程则与实际采集得到的规程的压下量比较接近, 负荷分配较为合理, 有利于保证最终的板形。

## 4 结论

本文根据中厚板生产产品规格变化大、轧制

条件复杂的实际情况, 探索研究将 CBR 与 KDD 技术相结合来进行负荷分配的方法, 提出了一种基于 CBR 和 KDD 的负荷分配模型, 合理地解决了这一特定问题中知识获取的问题, 充分利用了工人的操作经验, 提高了求解效率和求解质量, 使负荷分配更符合实际轧制情况。当然, 这种模型的应用效果在很大程度上要取决于案例数据的好坏, 将该模型用于实际生产要涉及案例数据的处理工作, 但这种数据处理的工作已较传统的方法(如 PH 图法)大为简化了。

## 参考文献

- [1] 孙本荣, 王有铭, 陈瑛. 中厚钢板生产. 北京: 冶金工业出版社, 1993
- [2] 鞍钢新轧钢股份有限公司厚板厂过程控制计算机系统改造初步设计书. 北京: 高效轧制国家工程研究中心, 1999
- [3] 田盛丰. 人工智能与知识工程. 北京: 中国铁道出版社, 1999
- [4] Leake D. CBR in Context: the Present and Future. Case-based Reasoning: Experiences, Lessons & Future Directions. AAAI Press, 1996
- [5] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large database. *Proc ACM SIGMOD Int Conf Manage Data*, 1993(2): 207
- [6] Agrawal R, Srikant R. Mining quantitative association rules in large relational tables. *Proc ACM SIGMOD Int Conf Manage Data*, 1996(2): 12
- [7] Agrawal R, Gupta A, Sarawagi S. Modeling multidimensional databases. In: *Proc. the 13th Int Conf on Data Engineering*. Birmingham, 1997

## Application of CBR and KDD techniques to modeling plate rolling schedules

Ma Dangxian<sup>1)</sup>, MU Zhichun<sup>2)</sup>, WANG Shaobo<sup>2)</sup>, WANG Changsong<sup>1)</sup>

1) Mechanical Engineering School, University of Science and Technology Beijing, Beijing 100083, China

2) Information Engineering School, University of Science and Technology Beijing, Beijing 100083, China

**ABSTRACT** In order to develop an adaptive model for automatically generating rolling schedules that comply with practical operations in heavy plate rolling, the techniques of Case-Based Reasoning (CBR) and Knowledge Discovery in Database (KDD) are introduced in the modeling process. In the model CBR not only is used for storing and retrieving the cases, i.e. schedules by which high-quality products have been rolled, but also generates an initial schedule for the material to be rolled according to the similarity in attributes between the material and the stored cases. KDD is introduced to find the rules from the operational data records to modify the initial schedule if there is difference between the ongoing rolling practice and the case applied. The experimental results of comparing the schedules against the practical rolling data show that the schedule generated by the new model is more reasonable and conformable to the actual rolling practice than that generated by traditional methods, which indicates that the method proposed in this paper is a promising method for rolling schedule generation modeling.

**KEY WORDS** rolling schedule model; case-based reasoning; knowledge discovery in database