

基于 Hadoop 的产品制造过程的碳足迹核算研究

江绪宇¹,李仁旺^{1,2},方旭斌¹,陈昆昌¹,贾江鸣¹

(1. 浙江理工大学机械与自动控制学院,杭州 310018;2. 浙江科技学院机械与汽车工程学院,杭州 310023)

摘 要:为准确核算产品制造过程中的碳足迹,首先采用 Hadoop 分布式存储技术存储产品加工数据,将碳足迹核算所需的产品实时加工数据进行预处理后,按照规定存储策略存储到分布式文件系统中,解决了碳足迹核算基础数据的来源问题。然后采用 MapReduce 分布式计算作为碳足迹核算的具体核算方式,根据产品各加工阶段的碳排放特性建立准确的碳排放核算模型,并将模型转化成 MapReduce 函数,快速获取制造过程的碳排放量。最后将所提理论应用到阶梯轴粗车阶段的碳足迹核算,结果表明所提理论可行。

关键词:碳足迹核算;HDFS;MapReduce;碳排放

中图分类号: X502

文献标志码: A

文章编号: 1673-3851(2018)03-0211-07

0 引 言

随着全球各个国家工业化进程的不断推进,大量二氧化碳的排放引发的各种环境问题逐渐显现出来。为避免人类生存环境进一步恶化,各国不约而同地将低碳发展、节能环保作为发展绿色经济的准则,越来越多的国家相继制定了碳排放法规,以减少那些能耗大的制造型企业的碳排放^[1]。碳足迹核算作为碳排放规章制度的衡量标准,对其研究具有重要的意义。

目前已有大量学者对产品制造过程碳足迹核算进行了研究。孙良峰等^[2]通过追踪零部件各个生产环节的碳排放,构建了复杂装备的碳足迹核算模型,但对于零部件的各个生产环节的碳排放并未给出相应核算方法。刘琼等^[3]根据产品制造过程与被加工产品之间的对应关系,确定产品制造过程的碳足迹总和,在碳足迹核算时,大多采用了车间的经验数据或者采取平均数据。Fang 等^[4]研究了设备切削速度、产品加工时间与设备功率之间的关系建立碳排放核算模型,但是在核算碳排放时,只考虑了设备运转时的电能消耗,忽略了冷却液等其他设备的碳足

迹核算。张翠侠等^[5]在核算产品碳足迹时,以产品各加工工位为一个碳排放单元,假设某些生产条件成立,给出了每个加工工位的碳排放核算公式。以上产品制造过程碳足迹核算基本上都存在所需基础数据缺乏的问题,这导致核算结果存在一定的误差,同时在碳足迹核算时,大多没有给出具体的数据处理方法。

本文针对以上机械加工车间产品碳足迹核算问题,将采集到的大量实时加工数据,按照以产品自身为核心的存储策略存储到 HDFS(Hadoop 分布式文件系统)中,解决碳足迹核算基础数据缺乏问题;在碳足迹核算时,根据产品制造过程中的碳排放特性,利用 MapReduce 分布式计算进行数据提取和碳排放总量核算,核算出产品不同加工阶段的碳足迹,再将各阶段核算结果累加到一起,从而得到产品整个制造过程的碳排放。运用 Hadoop 技术对提高产品制造过程的碳足迹核算效率和准确度具有重要意义。

1 制造过程数据采集与预处理

1.1 数据采集

为准确测定产品制造过程的碳足迹,需要全面

收稿日期:2017-08-18 网络出版日期:2017-11-09

基金项目:国家自然科学基金项目(51475434);浙江省自然科学基金项目(LY14G010007);浙江省食品物流装备技术研究重点实验室开放基金项目(KF2016003zd)

作者简介:江绪宇(1992-),男,皖六安人,硕士研究生,主要从事制造业信息化方面的研究。

通信作者:李仁旺,E-mail:renwangli@zstu.edu.cn

采集车间实时加工数据,这些数据包括机器的运行数据、加工工艺信息、加工产品自身数据等。将需要采集的数据分为机器运行数据、加工工序信息、产品自身信息、其他信息四大类型,确定其采集周期为 f 。所要采集的主要数据项如图1所示。

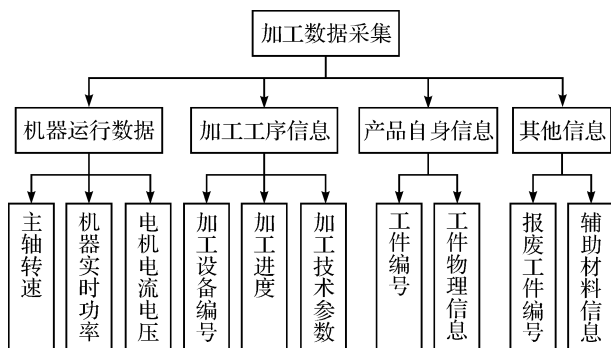


图1 主要采集数据

1.2 数据预处理

因数据直接来源于产品制造过程,实际生产环境、机器设备性能的差异,无法保证数据的洁净度以便于后期测定碳足迹。采集到的数据可能是不准确、含噪声的,这些异常数据不仅占用存储空间,而且对后期碳足迹测定造成不同程度的影响。因此在将数据存储到HDFS之前,需要对数据进行预处理^[6]。通过该步骤,可以使存储到HDFS中的数据符合规范,提高存储效率,同时也为后期碳排放核算做好准备。预处理工作主要有以下两步:

a) 数据清理

数据清理主要是发现采集数据中的一些异常数据,这些问题数据包含空缺值、噪声数据等。发现此类数据之后,可以通过修补、移除等方法来进行处理。通常数据处理包含定义错误类型、搜索错误数据、纠正错误数据、记录错误点和修改结果五个步骤。

b) 数据归约

数据归约的目的主要是为了尽量减少存储空间和保证数据的完整性,同时使用删除冗余或聚类等技术来压缩数据。例如在此阶段完成按照存储策略对数据的打包处理。

根据图1概括的采集数据项,并结合上述预处理技术,可将采集的数据项分为实时数据及固定数据。其中机器运行数据是实时数据,该类数据是时刻变化的,数据量较大,其他三大类数据基本为固定数据,不需要对其进行预处理。因此在数据预处理时主要考虑实时变化的数据,实时数据及固定数据的预处理流程如图2所示。

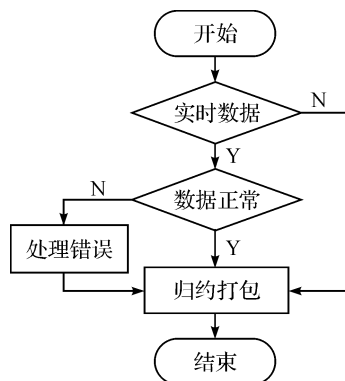


图2 数据预处理流程

2 数据存储与计算

2.1 存储策略

在产品制造过程中,采集到的加工数据类型繁杂且数据量庞大,在将数据存储到HDFS时需要对数据进行分类整理并建立合理的存储策略。结合产品制造过程碳排放的特性,建立了以产品为核心的存储模型。规定每个被加工产品在加工前会优先生成一个唯一ID,在该产品全加工周期中所有采集到的数据都会关联该ID,这样不仅能合理存储数据,也能保证在碳足迹核算时按照产品ID进行数据提取。本文综合以上情况确定数据的存储策略:

a) 所有数据都汇总到文本文件中存储,例如txt格式的文件,数据记录时间需要具体到秒;

b) 机器实时运行数据存储到一个文件,文件中每条数据的格式为:产品ID-机器ID-详细数据-记录时间;

c) 产品加工工序信息存储到一个文件中,文件中每条记录格式为:产品ID-工位信息-工艺参数-记录时间;

d) 产品自身信息存储到一个文件夹中,文件中每条记录格式为:产品ID-产品物理信息-流水信息-记录时间;

e) 其他基础信息存储到一个文件,文件中每条记录格式为:产品ID-详细数据-记录时间。

2.2 数据存储

由于产品制造过程数据的不间断采集,最终产生的数据规模比较大,因此将采集数据存储到HDFS中。HDFS是一种主/从模式的存储系统结构,可以由单台服务器扩展到上千台服务器^[7],其系统结构如图3所示。

NameNode(名称节点)服务器负责管理和维护HDFS目录系统并控制文件的读写操作,它作为整

个存储系统的主机对存储的数据节点进行管理,负责对文件系统树等元数据信息进行维护和管理并提供统一的文件命名空间。DataNode(数据节点)服务器负责存储数据。在 HDFS 中,文件是分块存储的,这些数据块被分散存储在不同的 DataNode 节点上,每个数据块又可以复制两份或多份。

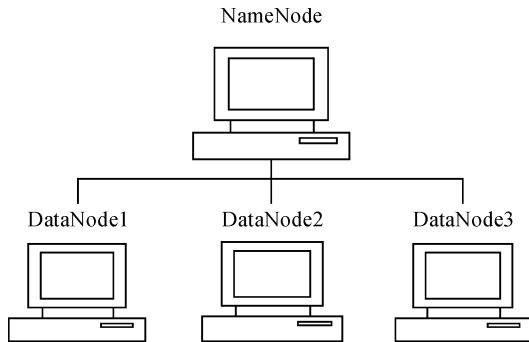


图 3 HDFS 分布式文件系统的结构示意图

2.3 MapReduce 分布式计算

在碳足迹核算模型的基础上,为提高大量生产数据的处理效率,采用 MapReduce 编程模型^[8]。MapReduce 属于 Hadoop 的分布式计算框架,当 Client 发起任务请求时,主节点 JobTracker 负责 Map、Reduce 任务的调度及控制,从节点 TaskTracker 负责任务的执行。MapReduce 有两个阶段:Map 阶段和 Reduce 阶段。两个阶段都以键值对形式作为输入和输出。Map 阶段将任务分割成多个更小的任务,由多个服务器阶段分别处理,此阶段的输入是经过预处理的数据。Map 阶段完成后,将任务结果汇总到 reduce 函数进行处理。MapReduce 运行示意图如图 4 所示。

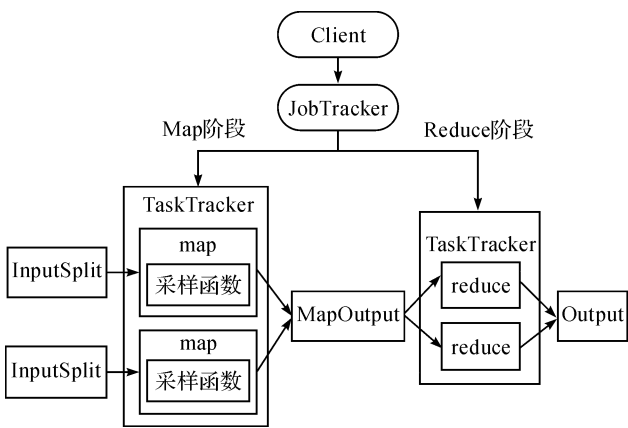


图 4 MapReduce 运行示意图

3 案 例

目前产品碳足迹核算边界有四种^[9],分别为从摇篮到坟墓、从摇篮到大门、从大门到大门、从大门

到坟墓。本文讨论的碳足迹是在制造过程中的碳足迹,属于“大门到大门”系统边界。依据物料衡算法^[10],产品制造过程碳排放核算需要确定各个系统消耗的各种能源,然后分别乘以对应的碳排放因子,最后累加即可得到总的碳排放量,常用资源的碳排放因子^[11]见表 1。

表 1 各资源碳排放因子

资源类型	碳排放因子	单位
电能	0.540	kgCO ₂ /(kw·h)
冷却液	5.143	kgCO ₂ /L
钢	3.220	kgCO ₂ /kg

3.1 基础数据存储

按照 2.1 节确定的存储策略,根据 HDFS 的存储原理,在将采集数据保存到 txt 文件中后,需要不断将最新的 txt 文件存储到 HDFS 文件系统中,数据存储的流程如图 5 所示。HDFS 数据存储首先由 HDFS 提供的客户端开发库 Client 定时发起存储请求,当确认请求成功时,NameNode 会建立或刷新关于 txt 文件的元数据并为 Client 分配对应的 DataNode,然后开始以流的方式写入第一个 DataNode,该 DataNode 写入完成之后,再将数据传递给 PipeLine 中的下一个 DataNode,直至最后一个 DataNode 完成存储。该过程涉及到 DistributeFileSystem、FSDataOutputStream 等程序类的操作。在存储容量上,HDFS 能够将多台电脑通过集群搭建在一起,从而实现存储容量的平滑扩容。

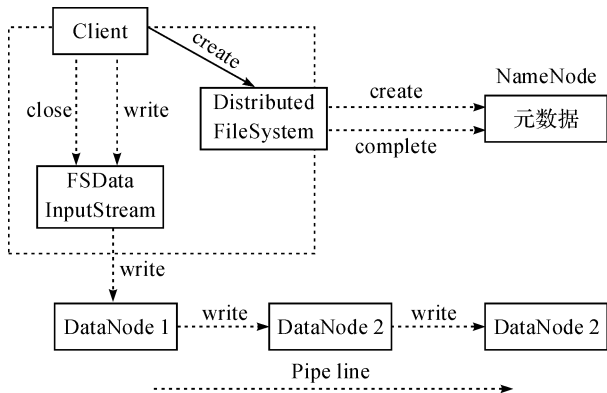


图 5 数据存储流程

3.2 建立碳足迹核算模型

以轴类加工车间阶梯轴加工为例量化分析其粗车加工阶段的碳足迹,该批次轴总数为 n 。根据轴加工实际碳排放特性,确定需要考虑的几种碳排放源,主要包含机床运转电能的碳排放 $W_{lathe,i}$ 、工件去除原材料的碳排放 $W_{mass,i}$ 以及生产辅助材料的碳排放 $W_{adit,i}$ 。因此轴在粗车阶段的碳排放 W_p 为:

$$W_p = \sum_{i=1}^n (W_{\text{lathe},i} + W_{\text{mass},i} + W_{\text{adit},i}) \quad (1)$$

3.2.1 机床电能的碳足迹核算

机床电能的碳足迹核算主要分为机床主传动系统的碳足迹和机床辅助系统的碳足迹,辅助系统的碳足迹核算主要是喷切削液以及灯光照明。该部分碳排放量为:

$$W_{\text{assist},i} = \alpha_e (P_1 \cdot t_1 + P_2 \cdot t_2) \quad (2)$$

其中: P_1 、 P_2 分别为冷却液泵和照明灯的功率, t_1 、 t_2 分别为冷却液泵和照明灯开启的时间, α_e 为电能的碳排放因子。

下面分析机床主传动系统的碳排放量。如图6所示,根据机床能耗特点可获得机床主传动系统输入功率变化折线图。根据图6可将主传动系统的碳排放量分为机床启动预热部分、机床切削加工部分、机床空载部分。

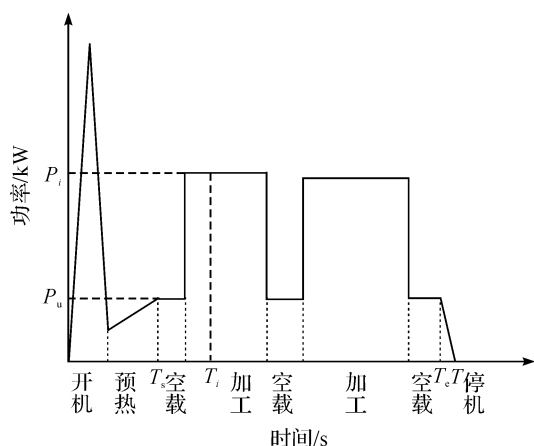


图6 设备运转功率折线图

a) 机床启动预热部分

机床启动是一个非常复杂的过程,功率波动很大,核算结果有一定偏差,但本次核算是一批轴加工的碳排放,只需计入一次启动过程,所以对最后结果影响不大。核算时需要先确定机床启动边界以确定启动时长,启动边界可确定为机床转速从零到稳定在目标转速的过程^[12]。在该启动时长上对功率进行积分即可算得开机预热的碳排放为:

$$W_{\text{start}} = \alpha_e \int_0^{T_s} P_i dt \quad (3)$$

其中: P_i 为 T_i 时刻的机床输入功率, T_s 为启动时长。由于数据采集周期为 f ,其值足够小。所以在核算启动过程碳排放时,可以采用离散求和的方式算得碳排放量,其原理如图7所示。对功率在 $0 \sim t$ 区间的积分可转化为 n 个区间相等且数值为 f 的长方形面积的和。

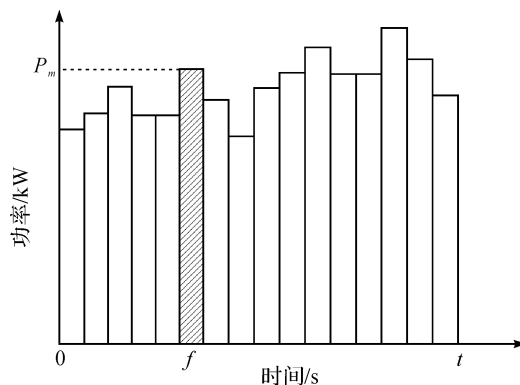


图7 离散求和原理图

根据离散求和的原理,对机床功率再 0 到 T_s 区间的积分即可转化为将采集到的离散功率点求和再乘以机床启动总消耗时间 T_s ,所以式(5)可转换成:

$$W_{\text{start}} = \alpha_e T_s \sum_{m=0}^n P_m \quad (4)$$

其中: n 为启动阶段总共采集到 n 个功率值, P_m 为第 m 个离散点的功率值。

b) 机床切削加工部分

阶梯轴粗车过程的碳排放量,除去开机预热过程,即为单个产品的粗车制造过程,对该过程求积分可得产品 i 的碳排放为:

$$W_{\text{product},i} = \alpha_e \int_{T_s}^{T_e} P_i dt \quad (5)$$

其中: T_s 、 T_e 为粗车产品的开始和结束时间。参考离散求和对式(3)的处理,所以式(5)可转换为:

$$W_{\text{product},i} = \alpha_e (T_e - T_s) \sum_{m=0}^n P_m \quad (6)$$

c) 机床空载部分

机床空载有两种情况,分别为工件切换过程的空载和加工同个工件调整刀具位置的空载。调整刀具位置的空载产生的碳排放已经包含在机床粗车轴部分,因此该部分只需考虑工件切换过程的空载。需要确定的数据为机床空载功率 P_u 和切换工件所需时间。空载过程中机床主轴不受力,所以 P_u 基本为定值,切换工件所需时间跟工人操作熟练度有关,该值可以根据前后两个阶梯轴的加工结束时间和加工开始时间相减得到。

$$W_{\text{idle},i} = \alpha_e P_u (T_{s,i+1} - T_{e,i}) \quad (7)$$

其中: P_u 为机床空载功率, $T_{e,i}$ 为第 i 各阶梯轴的加工结束时间, $T_{s,i+1}$ 为第 $i+1$ 个阶梯轴的加工开始时间。当 $i=n$ 时, $W_{\text{idle},i}$ 为零。

3.2.2 工件材料损耗的碳排放量

工件原材料消耗量可以通过去除材料的质量进行核算,则生产该阶梯轴原材料产生的碳排放核算公式为:

$$W_{\text{mass},i} = (m_1 - m_2) \alpha_m \quad (8)$$

其中: m_1 为工件 i 加工前质量, m_2 为工件 i 加工后质量; α_m 为加工轴材料的碳排放系数。

3.2.3 冷却液的碳排放量

加工车间内目前没有冷却这些消耗品的监测系统,同时这些消耗品使用周期较长。所以在核算时,可以根据车床使用规范,获得这些消耗品一个周期的时间及更换一次的容量,再平均到每个产品上。

$$W_{\text{adit},i} = \alpha_1 \times \left[\frac{L}{T_1} (T_e - T_s) \right] \quad (9)$$

其中: α_1 为冷却液的碳排放因子, L 为更换一次冷却液

的总量, T_1 为一个使用周期的时间长度,以秒为单位。

将式(2)、(4)、(6)~(9)代入式(1)中即可得粗车阶段该批次阶梯轴总的碳排放量:

$$W_p = W_{\text{start}} + \sum_{i=1}^n (W_{\text{product},i} + W_{\text{idle},i} + W_{\text{assist},i} + W_{\text{mass},i} + W_{\text{adit},i}) \quad (10)$$

3.3 碳排放核算

运用 MapReduce 计算同批次阶梯轴总的碳排放量时需要经过两个阶段,分别为数据提取阶段和碳排放核算阶段,因此建立两个 MapReduce 程序,碳排放核算数据流如图 8 所示。

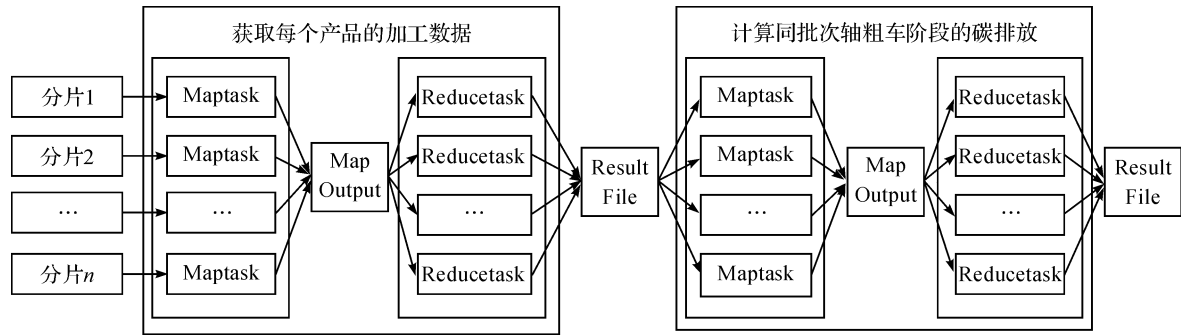


图 8 碳排放核算数据流

在第一个 MapReduce 程序中, Map 程序根据阶梯轴的 ID 提取各个分片中该阶梯轴的相关加工数据并得到 output 文件,各阶梯轴的加工数据会按照 ID 排序依次存储在 output 文件中。当 Map 阶段完成之后, JobTracker 根据 output 文件内容分配任务给各个 reduce 函数, reduce 函数会将同一个阶梯轴的加工数据汇总在一起。第二个 MapReduce 程序用于核算同批次轴粗车阶段的碳排放总量,该阶段首先会根据第一阶段的结果文件, JobTracker

对其进行分片后分配给 map 函数进行计算,该 map 函数会计算各个阶梯轴粗车阶段的碳排放总量并得到 output 文件,再由 JobTracker 根据阶梯轴 ID 分配一定数量的单个阶梯轴碳排放结果给各个 reduce 函数,由 reduce 函数计算得出同批次阶梯轴总的碳排放量。依据 3.2 节所述阶梯轴碳足迹核算模型以及碳排放核算数据流,编写代码进行粗车阶段的碳足迹核算,即可得出同批次阶梯轴的碳排放总量。MapReduce 进行碳排放核算的核心代码块如下:

```
//获取机床辅助设备的碳排放量
int getAssit(int elefactor, List<Integer> assitPowList, List<Integer> assitTimeList){
    int assitresult=0;
    for (int i=0; i<assitPowList.size(); i++){
        assitresult += assitPowList.get(i) * assitTimeList.get(i);
    }
    return assitresult * elefactor;
}

//机床运行过程中的碳排放量
int getLathe(int elefactor, List<Integer> startList, List<Integer> productList, List<Integer> idleList) {
    int latherresult=0;
    for (int i=0; i<startList.size(); i++) {
        atherresult += startList.get(i) * startTime;
    }
}
```

```

    for (int i=0; i < productList.size() ; i++) {
        latheresult += productList.get(i) * productTime;
    }
    for (int i=0; i < idleList.size() ; i++) {
        latheresult += idleList.get(i) * idleTime;
    }
    return latheresult * elefactor;
}
//轴材料损耗产生的碳排放量
int getMass(int elefactor,int prequility,int aftqulity) {
    return elefactor * (prequility-aftqulity);
}
//冷却液的碳排放量
int getAdit(int aditfactor,int capcity,int cycleTime) {
    return (capcity/cycleTime) * productTime * aditfactor;
}

```

由于碳足迹核算所用数据包含了零件质量等物理数据、冷却液使用量及周期等数据,可以准确得出工件去除原材料的碳排放以及生产辅助材料的碳排放;同时包含了机床加工过程中的实时功率等数据,并使用离散求和的方式得出机床运转电能的碳排放,考虑了加工过程中的机床功率波动情况。相比文献[2—5]使用经验公式得出机床功率或采取平均算法等其他不精确的核算方法,本文提出的方法可以更精确地核算出产品制造过程的碳足迹。在批量产品的碳排放核算时,按产品 ID 分割 Map 任务同时核算并将结果汇总到一起,即可得出粗车阶段的总碳排放量,实现了碳排放量的自动核算,同时 MapReduce 并行计算技术使计算效率得到了很大的提高。

4 结 语

本文结合产品制造过程碳足迹核算的研究现状,提出将产品制造过程中的生产数据存储在 Hadoop 分布式文件系统中,并将存储的实时加工数据运用到碳足迹核算中,提高了碳足迹核算的准确度。利用 Hadoop 集群技术及其分布式计算核算产品制造过程的碳排放,将碳足迹核算模型转变成具体代码,给出了具体的数据计算方法。最后以阶梯轴加工为例,结合制造过程的碳排放特性给出了碳足迹核算模型,依据碳足迹核算模型编写程序,再提取具体加工过程数据传入程序中得出粗车阶段的碳排放量,在大数据量的提前下也能保证计算速度。由于本文只是讨论碳足迹核算与 Hadoop 结合的可

行性,并未对核算结果进行直观展示,在后面的研究中,可将核算结果存储到 MES 系统中,实现整个车间的碳足迹监控。

参考文献:

- [1] 东艳. 全球气候变化博弈中的碳边界调节措施研究[J]. 世界经济与政治, 2010(7): 65-82.
- [2] 孙良峰, 裴乐森, 张树有, 等. 面向低碳化设计的复杂装备碳排放分层递阶模型[J]. 计算机集成制造系统, 2012, 18(11): 2381-2390.
- [3] 刘琼, 田友全, 周迎东, 等. 产品制造过程碳足迹核算及其优化问题[J]. 中国机械工程, 2015, 26(17): 2336-2343.
- [4] Fang K, Uhan N, Zhao F, et al. A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction[J]. Journal of Manufacturing Systems, 2012, 30(4): 234-240.
- [5] 张翠侠, 刘从虎, 张明玉, 等. 机械加工车间碳足迹监测与预警方法及关键技术[J]. 计算机集成制造系统, 2015, 21(10): 2677-2686.
- [6] 刘莉, 徐玉生, 马志新. 数据挖掘中数据预处理技术综述[J]. 甘肃科学学报, 2003, 15(41): 117-119.
- [7] 栾景超, 马志强, 李昊甦, 等. Hadoop 分布式文件系统资源管理器的设计与实现[J]. 科研信息化技术与应用, 2014, 5(1): 41-52.
- [8] 杨震宇. 基于 MapReduce 框架下的数据挖掘方法研究[J]. 中国高新技术企业, 2017(4): 8-10.
- [9] Spatari S, Betz M, Florin H, et al. Using GaBi 3 to perform life cycle assessment and life cycle engineering[J]. The International Journal of Life Cycle Assessment, 2001,

- 6(2):81-84.
- [10] 张肖,向晓东,刘汉杰,等. 钢铁行业碳排放量核算方法的实证性研究[J]. 工业安全与环保,2012,38(6):86-88.
- [11] 梁国熙,梁耀彰,陈海寿,等. 香港中小企业碳审计工具箱[EB/OL]. (2010-02-01)[2017-08-18]. http://www6.cityu.edu.hk/aerc/sme/images/sme_chi.pdf.
- [12] 刘飞,刘霜. 机床服役过程机电主传动系统的时段能量模型[J]. 机械工程学报,2012,48(21):132-140.

Research on the carbon footprint accounting of product manufacturing process based on Hadoop

JIANG Xuyu¹, LI Renwang^{1,2}, FANG Xubin¹, CHEN Kunchang¹, JIA Jiangming¹

(1. Faculty of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China;

2. Faculty of Mechanical & Automotive Engineering, Zhejiang University of Science and Technology, Hangzhou 310023)

Abstract: In order to calculate the carbon footprint of the product manufacturing process accurately, the Hadoop distributed storage technology was firstly applied to store product processing data. After pre-processing of product real-time processing data required by carbon footprint accounting, the data were stored in the distributed file system according to the required storage strategy, which solved the basic data source problem of carbon footprint accounting. Then, MapReduce distributed computing was used as a method of calculating carbon footprint. Accurate carbon emission accounting model was established according to the carbon emission characteristics of each product processing stage, and then the model was converted to MapReduce function to rapidly get carbon emission of manufacturing process. Finally, the theory was applied to calculate carbon footprint in the rough turning stage of stepped shaft, which verified feasibility of the proposed method.

Key words: carbon footprint accounting; HDFS; MapReduce; carbon emission

(责任编辑:康 锋)