

基于回溯剪枝算法的燃烧反应路径可视化分析

郑瑞林, 陈志龙, 王 晶

(武昌工学院, 湖北 武汉 430000)

摘要:为响应国家节能减排改善环境质量的号召,我国汽车产业节能降耗势在必行,提高内燃机燃烧效率,降低尾气有害物质排放是其中强有力的手段,目前针对内燃机内部燃烧反应的研究多是基于反应分子动力学用大规模分子并行模拟器(LAMMPS)进行燃烧模拟实验,但由于内燃机内部燃烧反应是极为复杂的链式反应,后期实验数据非常庞大,且其中有大量的重复数据和异常数据。人工进行数据清洗并统计结果要消耗大量的时间和精力,错误率还高,严重耽误实验进程。本项目的开发即是希望通过计算机代替人工,用程序实现内燃机内部燃烧反应后期数据清洗的自动化,并运用回溯剪枝算法处理技术和数据可视化技术,将原本繁杂抽象的数据以图表的形式直观显示,简化人工操作,在缩短数据处理周期的同时降低错误率,对推进内燃机内部燃烧反应的相关研究有重要意义。

关键词:内燃机;燃烧反应;数据清洗;回溯剪枝算法;数据可视化

中图分类号:TP309

文献标识码:A

文章编号:2096-9759(2023)06-0126-05

1 引言

随着国内经济的迅速发展,民用汽车保有量日渐增加,汽车尾气所带来的大气污染问题也日益严重,要解决汽车尾气对环境的污染问题,最直接的办法是对尾气产生的源头内燃机燃烧反应进行改进,提高内燃机燃烧效率,尽可能使燃料完全燃烧(完全燃烧时只产生水和二氧化碳),最大程度释放热量,尽量减少有害颗粒物的排放。但内燃机燃烧反应的整个过程极为复杂,用大规模分子并行模拟器处理后实验数据量非常庞大,数据的高效处理一直是内燃机燃烧反应研究中难以解决的一个问题。

且从根本上来说内燃机的燃烧反应也是一种化学反应,遵循化学反应动力学的基本原理,根据苏联科学家谢苗诺夫提出的链式反应理论,物质燃烧经历以下几个过程,首先可燃物和助燃物质需要吸收能量,分解为游离态的原子或离子,游离态的原子或离子与其它分子反应,所产生的产物又继续反应,形成一系列链式连锁反应,进而放出热量。这些游离态的原子或离子也叫游离基,因为数量非常庞大,不仅会对整个燃烧反应方向造成一定的影响,造成理论产物与实际产物的部分偏差,还会在实验过程中产生大量繁杂的实验数据难以处理。

目前要研究这种数据量庞大的分支链式反应,国际上比较主流的方法是反应分子动力学(ReaxFF MD)。但反应分子动力学的工具 LAMMPS 没有对异常数据和重复数据的去除功能,导致 LAMMPS 模拟内燃机燃烧反应后所产生的结果数据异常庞大,之前完全靠人工处理,整个过程将花费大量的时间和精力,错误率还高。导致在研究内燃机燃烧反应整个过程中,大量的时间浪费在了数据清洗处理上,严重耽误了研究进程。

于是,本项目以内燃机燃烧反应研究项目组所提供的燃烧反应数据为基础。针对内燃机内部燃烧反应研究统计数据难,分析数据难,错误率高,耗时长的问题。将数据处理自动化,由计算机自动识别数据清洗去重,用算法对可逆链式反应进行自动遍历剪裁,简化了人工操作,大幅降低了出错率,对

处理后的数据用 Python 的 PyEcharts 包进行可视化处理,让结果数据图像化,直观化,整个处理过程相较于人工,效率得到明显提升,正确率大幅度增长,极大的加快了内燃机内部燃烧反应的实验进程。

2 项目构建总体思想概述

本项目根据链式反应理论,以内燃机燃烧反应为研究基础,将原来需要人工统计的大量内燃机燃烧反应数据,用计算机程序实现自动化统计。旨在用计算机替代人工,提高数据统计工作效率,且在数据统计完成之后,将数据注入到数据库中,用数据清洗程序和回溯剪枝算法进行数据自动化清洗,对不符合条件的重复数据和可逆反应自动识别去除,避免了人工在面对大量数据时出现遗漏和误除。在数据清洗结束后,将通过 Python 程序从数据库中进行数据自动化读取,不同图表对数据信息的需求不同,在数据库中读取时要分类处理,逐步读取,最终调用 PyEcharts 包进行结果数据的可视化处理,生成数据链路图,分枝树状图,效率矩形图等多个可视化图表,使反应链式关系更直观。具体技术路线如图 1 所示。

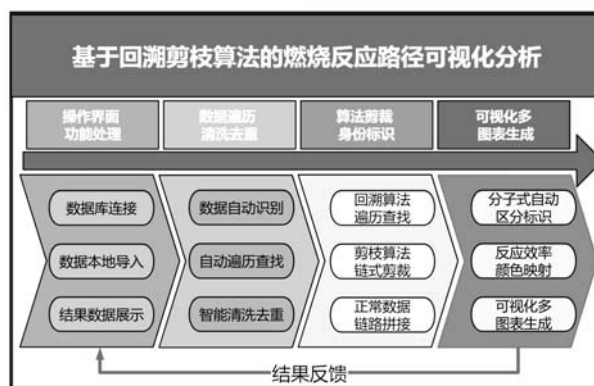


图 1 技术路线图

3 项目功能整体概述

软件整体操作界面用 C#语言进行设计,程序主体用 Py-

收稿日期:2023-02-16

基金项目:国家级大学生创新训练计划项目(202213241006);武昌工院校级科研项目(2021KY03)。

作者简介:郑瑞林(2000-),男,四川成都人,本科生,学士,主要研究方向:数据分析及可视化;陈志龙(2001-),男,浙江温州人,本科生,学士,主要研究方向:数据分析及可视化;王晶(1985-),女,山东济南人,硕士,副教授,主要研究方向:数据挖掘,数据分析及可视化。

thon 进行数据处理,并调用 Python 中的 PyEcharts 包进行图表可视化处理,但 C#对 Python 的支持度并不高,虽然有 IronPython 可以使用,且该语言与 .NET Framework 紧密集成,但是 IronPython 对 Python 的第三方库并不支持,如果调用 Numerical Python 和 Python Data Analysis Library 会出现无法找到的问题,本项目针对此问题,将 Python 程序打包为 EXE 文件,C#再对 EXE 文件进行调用,从而实现 C#和 Python 的交互,且不会出现第三方库无法找到的问题。软件操作界面如图 2 所示。



图 2 软件操作界面图

(1) 数据库状态模块

设有数据库账号,数据库密码,数据库端口输入框,因为后期数据导入后在数据预处理部分是需要将数据注入本地数据库进行处理的。所以在使用前,用户要输入本电脑的数据库账号和密码,数据库端口用户可以自由选择,在安装时默认为 3306,如果安装时修改了端口则需要修改端口号。用户在输入完成后点击登录按钮,连接成功时输出界面会输出 success 字符,提示登录成功。

(2) 文件路径选择模块

对于文件路径选择模块,用户可以选择数据导入按钮进行源数据的导入,用户在导入时只用选择到源数据的上一层文件夹,点击导入即可,为了使用的方便性,在导入框的下面设置了记住路径按钮,用户在下次使用时就不用再次选择数据源路径,可直接使用。

(3) 软件重要参数选择模块

为了满足不同的实验要求,在一些特殊条件下可能要对去除的重复数据和异常数据进行另存保留,以便后期查询,所以提供了是否另存去除的异常数据和是否另存去除的可逆反应的选项,供研究人员根据不同情况进行选择。

(4) 程序处理步骤模块

在程序处理步骤模块,有 LAMMPS 源数据处理选择框,重复数据清洗处理选择框,异常数据清洗处理选择框,可逆反应清洗处理选择框,数据可视化预处理选择框,用户可以根据数据的实际情况和实际的需要进行处理步骤的选择,不同的处理步骤的选择,会导致最后的结果出现不同的差异。

(5) 可视化图表选择模块

在这个模块,用户可以选择分枝树状图,数据链路图,效率矩形图,每张图对应不同的结果展示,用户要分析效率就选择效率矩形图,效率矩形图中矩形面积大小代表反应效率大小。要进行内燃机燃烧反应支链分析就选择分枝树状图,分

枝树状图会将内燃机燃烧反应整个支链中的各个结点以及结点间的连接情况直观的展现出来。

4 项目各模块实现原理概述

4.1 反应物数据自动清洗实现模块

数据清洗处理是对数据进行可视化展示的必要过程,没有经过筛选和记录的无关联性数据对实验研究没有任何作用。因为内燃机内部燃烧反应的特殊性,用 LAMMPS 对内燃机内部燃烧反应进行轨迹模拟,最后得到的结果只是化学表达式标注的分子片段随时间变动的客观规律,并不能获得反应轨迹模拟过程中所产生的具体化学信息。导致最后得到的模拟结果数据较为混乱,辨识度较差,无法直观的观察出内燃机内部燃烧反应的具体化学反应链走向。需要后期人工辨识具体初始反应物,以及反应过程中产生的各种复杂的中间物质,并对各个支链反应进行拆分处理,使其产生链式关系并记录相应反应路径的反应次数。

因为内燃机内部燃烧反应是极为复杂的分支链式反应,其反应产生的游离态的原子或离子比普通分子平均动能具有更高的活化能,造成这些原子和离子异常活泼,所以这些游离态的原子或离子又称为中间活性产物,这些中间活性产物极易和其他物质的分子进行反应进而产生新的物质或新的中间活性产物,如果这个过程中没有产生稳定的分子,整个过程将持续往复进行下去,直到所有的中间活性产物转变为稳定的分子,连锁反应中断,整个反应结束。整个反应过程中将产生大量的分支线路,进而产生大量的分支产物,但不是所有的分支产物都对研究有用,很大一部分分支产物是对研究没有意义的,在大规模分子并行模拟器中具体体现就是产生大量的重复数据和无效数据,之前需要由人工分析去除,不仅耗时耗力,还由于人工精力有限,难免会出现漏除,误除的情况,导致最后的结果不准确。

针对这些问题,本程序以人工去除方法标准为基础理论,将燃烧反应数据清洗步骤向自动化方向改进,实现重复数据和无效数据自动识别去除。首先对 LAMMPS 单次模拟计算后的数据结果进行数据读取,LAMMPS 模拟的数据保存在 bond 文件中,其中包括了分子 id,原子 id,原子类型,成键数量,第一个键所连原子 id 等多个信息,且 LAMMPS 对化学元素根据元素周期表有默认的编码,例如默认将 C、H、O、N 元素编码为 1、2、3、4。但在研究过程中并不是所有的信息都要用到,只有一部分信息作为判断依据,为了方便后期的清洗去重,减少字符串遍历时间,则需要根据所需要的列表信息将原来的 bond 数据提取注入到数据库中,得到以分子 id 为主键的数据库表,且模型中的每一个原子都有一个 id 和计算机编码,就算是同一类原子,例如 C 原子,每一个 C 原子的 id 也都是不同的,bond 文件关键信息如图 3 所示。

C2H6 中的第一个碳原子原子 id 为 1,第二个碳原子原子 id 为 2, type 为 1 则表示该元素为 C, id 为 1 的碳原子共连接四个原子,原子 id 分别为 6、7、2、11。id 为 2 的碳原子共连接四个原子,原子 id 分别为 8、9、10、1。以此数字编号来唯一确定一个分子。

但也正因为分子数字编号的唯一性,两个完全相同的分子数字编号也是不同的,导致 LAMMPS 对出现的重复数据无法辨别,需要后期进行处理,对这样的重复数据,首先将一个分子反应后的所有产物放到一个列表里,注意不是一条链上

的分子,如果是一条链上的分子有重复叫可逆反应,在下文中会详细讲解,对一个分子反应后的所有产物用字符串的遍历,将遍历到的字符串,放到一个新的空列表中,然后将这个分子和其他分子做字符串的对比,如果有相同的就删除掉相同的分子只保留一个,整个过程循环进行,直到所有的重复数据删除完成。当然在这其中还有很多的无效数据,和现实情况差距较大,则在重复数据删除完成后,用正则表达式对异常数据辨别,发现超过正常值就进行删除,直到所有的数据都正常,重复数据和无效数据清洗操作流程如图4所示。

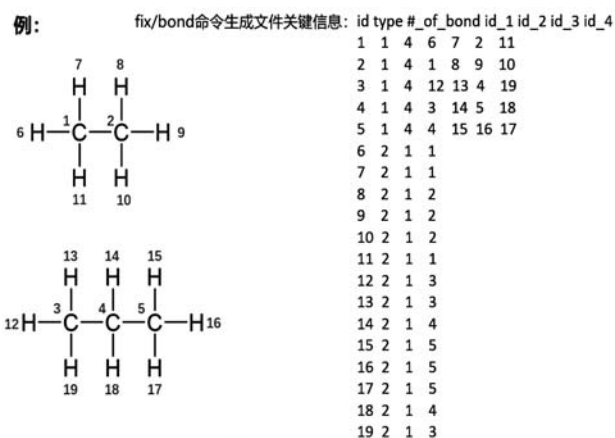


图3 bond 文件键信息图

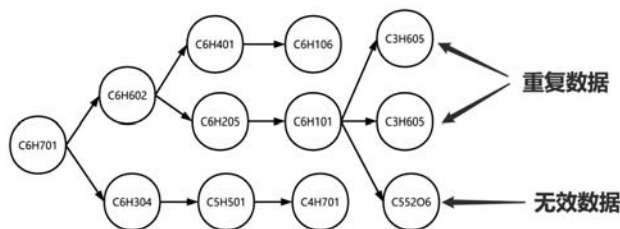


图4 重复数据和无效数据清洗操作流程

4.2 回溯剪枝算法对可逆反应的自动处理模块

回溯剪枝算法对可逆反应进行处理是数据清洗到数据可视化非常重要的一个环节,原因在于内燃机内部燃烧反应不仅是链式反应,还是链式反应中较为复杂的分支链反应,也就是多个中间活性产物和多个其他物质的分子进行反应,产生多个新的中间活性产物或多个其他物质,只要中间活性产物没有完全消失,就会一直进行下去,每一个产生的新物质的分子就是一条支链的新节点,这种产生是一种趋近于指数式的爆炸式增长,一种坏处是会导致大量的实验数据产生,研究统计起来非常困难,还有就是这种反应机制的庞大性和复杂性还会有几率产生可逆反应。也就是部分中间不稳定活性产物与其他物质分子进行多次反复反应,可能某一步反应产生的物质分子,和前面步骤的某一个物质分子相同了,也就是多次反应过后的产物和反应前的物质一样。这对于研究内燃机燃烧反应内部反应链变化和各个分支链效率区别是没有意义的。甚至因为部分可逆反应的存在,对后期统计数据时会造成影响,造成实验结果的较大偏差,是必须进行去除的。

之前数据清洗过程中无法去除可逆反应的原因主要是通过数据的遍历和对比去除的是单个或多个的异常数据和重复数据,而可逆反应是要去除整个可逆反应链,整个支链可能只有中间一小段或中间多个小段是可逆反应链,而两端是正常的,这时就不能对整个支链进行去除,只能对中间一个或多个

可逆反应链进行剪裁,对剩下的正常反应链进行拼接,操作流程如图5所示。

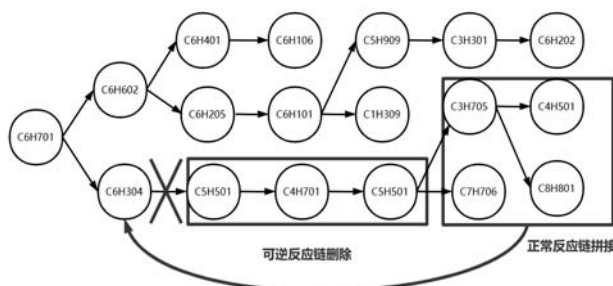


图5 可逆反应链剪裁拼接操作流程

整个过程普通的数据清洗过程无法实现,这时就需要用到回溯剪枝算法,回溯剪枝算法中的回溯和剪枝其实是分两个步骤来进行的,详细的名称应该叫回溯算法和剪枝算法。其中回溯算法类似于一种穷举或试探,是一种以实验为主,在不断试错后找到正确方法的算法,用程序语言来说就是在集合中进行递归处理,不断查找子集,最后集合本身的大小就是整个树的宽度,而整个递归进行的深度,就是整个树的深度。在本程序中回溯算法主要应用于可逆反应的搜索上,其特点是每一个产物都有若干个产物,每一个产物的反应过程可以分为若干个步骤,得到其所有反应物反应流程是一个不断尝试的过程。

在程序实现中,首先按照一定的顺序枚举反应物出现的每一个产物,之前已经出现的反应物在接下来要选择的反应物中不能出现。因为每一个反应产生的反应产物在下一个反应中又充当了反应物的角色,整个反应是链式进行的,一环扣一环,为了保证每个反应物的反应产物都被取到,需要采用for循环的方式,循环的次数是每一个反应物产生反应产物的数量。在这棵树的结构中,for循环解决了横向遍历问题,要这些数据成为可用的链式结构,还需要用递归进行纵向遍历。即从根节点出发,向符合条件的第一个子节点出发搜索,然后以当前子节点作为根节点继续搜索,如果找到符合条件的反应产物则记录下来,如果找不到则从当前子节点返回上一级父节点,从父节点的下一个子节点进行搜索。递归的终止条件是搜索到叶子节点,也就是生成的最终产物之一。如果发现可逆反应就储存结果,并结束本层递归。

另外,由于执行的是深度优先遍历,从较深层的结点返回到较浅层结点的时候,需要做“状态重置”,也就是回溯。因为一个反应物不只有一个反应产物,那么其反应子链也不止一条,整体成树枝状分布,那么每一个树枝的分叉点,就是一个搜索的回溯点,当一条子链搜索完毕后,不管有没有结果,都要退回到回溯点进行另一条子链的搜索,以保证每一条子链都被搜索到。当搜索完毕后找到了所有的可逆反应,这时就要进行剪枝处理,这里的剪枝指的是将有可逆反应的那一段进行删除,但是前后正常的部分要进行保留,并在可逆反应链删除后进行拼接。在程序中实现为将可逆反应链的最后一个反应物之后的数据包括反应物本身保留到一个空的列表中,而将可逆反应链第一个反应物之后的数据包括其本身全部删除,删除完成后将之前保留到空列表中的数据拼接或删除完数据的支链上,这时可逆反应链就被删除了,如此往复直到所有的可逆反应链被删除。整个过程总结起来就是for循环横向遍历,递归纵向遍历,回溯不断调整结果集,剪枝处理结果集。

4.3 结果数据自动可视化模块

为方便研究者快速、直观地认识内燃机燃烧反应的内部过程,对内燃机燃烧反应数据的可视化展示也是重要的步骤。结果数据的可视化是将繁杂的数据转换为直观视觉图形以有效传达燃烧效率、反应关系、反应趋势等信息的过程,是内燃机燃烧数据分析实现自动化的一个重要方面。虽然结果数据经过了数据清洗和回溯剪枝算法剪裁,数据量远远低于处理前,但静态繁杂的结果数据依然是抽象的,难以发现燃烧反应各支链间的联系和其中存在的问题,而结果数据可视化可以将静态繁杂的数据通过动态的图像直观展示,帮助研究人员快速、轻松地理解内燃机燃烧反应中复杂的信息。

在数据可视化中使用了 Python 可视化包 PyEcharts,载入经过数据清洗和数据剪裁后的结果数据,调用 PyEcharts 包中的接口,形成可视化图表。因为需要生成多个可视化图表,以便研究人员能够根据具体需求选择不同图表进行效果展示,而 Python 的 PyEcharts 包中不同的图表所需的数据信息不同。绘制效率矩形图需要提取结果数据中每一步反应的反应效率数据,反应效率数据之前在 LAMMPS 中已经自动生成,在结果数据库中查找到每一步反应的反应效率数据与效率矩形图的接口相匹配,然后才能自动化绘制效率矩形图。对于分枝树状图,由于内燃机燃烧反应的中间态具有层次性及方向性的特点,需要采用 Sugiyama 算法进行布局,以燃烧反应中每一步的反应物及产物为节点,将每一步反应的反应物和产物依照反应先后顺序进行连接,形成一张由反应物及反应产物为节点,反应关系为连接边的树状图,树状图中每一条分枝即是燃烧反应中的一条支链。以下是内燃机燃烧反应结果数据可视化处理的图表原理和图表应用的详细介绍。

效率矩形图以颜色区分不同的反应支链,以矩形面积大小代表不同的反应效率,矩形面积越大反应效率越高,且针对一些面积相差较小的矩形,用肉眼难以分辨,以矩形面积大小表示效率的同时,还在矩形中标识了具体面积大小的数字信息,矩形面积肉眼无明显差距时,可进行参考。依次点击面积最大的矩形即可找到反应效率最高,反应最理想的线路。且可以通过观察矩形相应的分子名称和效率数值判断化学反应中的主要路径,以此实现观察整个反应中占主导地位的反应链的链长以及其中相关反应物的总反应次数。不用再由人工按照数字对比每条支链的反应效率来寻找反应效率最高的线路,效率矩形图如图 6 所示。



图 6 效率矩形图

分枝树状图可以完整的展示内燃机燃烧反应中所有分子的化学反应网络,对整个反应链路每条支链及其子链的情况能直观的展示出来,需要对相应的支链进行查看,就直接点击那个支链的起始节点,整个支链就会显现出来,可以更为全面的了解到整个燃烧反应网络。为后期支链数据的查询,如找到反应效率最高的支链后,要查询其具体内部反应网络,提供很大方便。分枝树状图如图 7 所示。

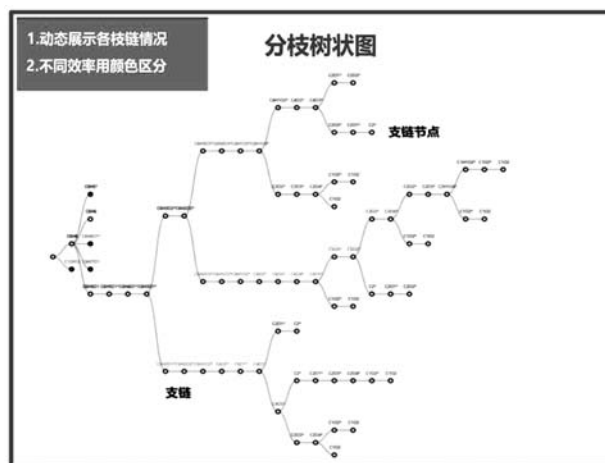


图 7 分枝树状图

数据链路图以碳原子数对反应链路进行分类,相同碳原子数的反应支链被聚集到一个界面进行展示,在碳原子数相同的反应支链中对不同类的反应物用颜色进行标识,通过调节可以看到同一类反应物的情况,在找到反应效率最高的线路后,若要研究与效率最高链路同碳原子数的支链线路与其他普通支链线路反应产物以及每个节点之间反应效率的具体区别,就可以用到数据链路图。数据链路图如图 8 所示。

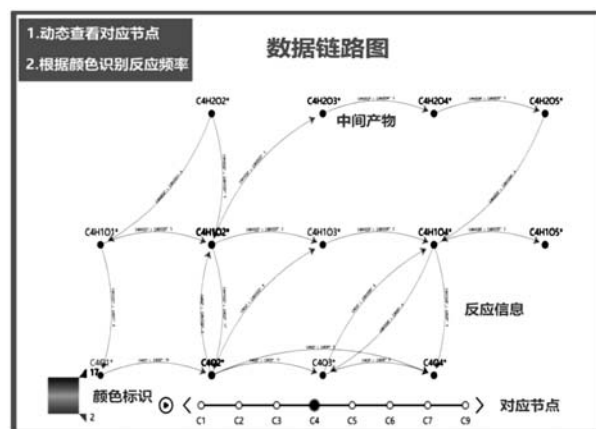


图 8 数据链路图

5 结语

本项目针对大规模分子并行模拟器模拟内燃机内部燃烧反应后期实验数据量庞大,异常数据人工难以处理的问题,通过以计算机代替人工的方式,将数据清洗自动化,用程序智能判断异常数据并进行去除,对人工难以去除的可逆反应用算法进行自动识别剪裁,最后对结果数据进行多图表可视化,让整个实验结果直观的展现出来。本项目以数据处理智能化自动化的方式解决了内燃机燃烧反应前期数据清洗效率低,

TYJL- II 型计算机联锁的结构与故障处理

王凤姣

(重庆公共运输职业学院, 重庆 402247)

摘要:TYJL- II 型计算机联锁系统是铁道部科学研究院研制开发的系列计算机联锁系统,具有完善的自我诊断、维护和远程诊断、与 CTC 系统结合、故障弱化、较强的防雷和抗电磁干扰性能。文章主要介绍了该系统的结构以及故障处理方法,并且选取了故障案例进行研究。

关键词:计算机联锁;故障处理;轨道交通

中图分类号:TP309

文献标识码:A

文章编号:2096-9759(2023)06-0130-03

1 引言

随着轨道交通运输的加速发展,原 6502 电气集中联锁系统由于在应用过程中存在维护成本高等不利因素,计算机联锁系统逐渐取代 6502 电气集中联锁系统的地位,在轨道交通运输当中广泛应用。计算机联锁系统是使用计算机进行实时控制的实例,它的主要特点有:实时性强,能在限定时间内对所控制的事件作出响应,并能随时中断;具有充分的过程输入和控制输出能力,能够从室外采集相关设备的状态,对采集结果进行处理以后将结果输出;具有通信和联网能力;高可靠性,计算机联锁系统能够在工业环境中长时间工作,并且大多采用冗余结构,具有更高的可靠性;可维护性,系统在设计的初始就考虑到后期的维护及可适应,计算机技术更新迭代快,在更新技术出现以后,不必更换整套系统,可在原系统上进行更新^[1]。

2 TYJL- II 型计算机联锁系统结构

TYJL- II 型计算机联锁是分布式多计算机系统,冗余方式为双机热备,其结构如图 1 所示。整个系统划分为三个层次,分别是人机会话层,联锁运算层和输入输出接口层。其中控制台和维修机采用的是单套配置,上位机、联锁机和执表机采用主备双机配置,并且双机采用的是双机热备方式,双机热备就是平时两台设备同时工作,在发生故障时可以自动进行主备切换,也可

通过人工进行手动切换,采用热备的优点是切换时用时短,备用设备切换以后可以直接参与工作,控制台只有一台,平时只与上位机主机进行信息交互,可以通过切换单元更改连接方式。

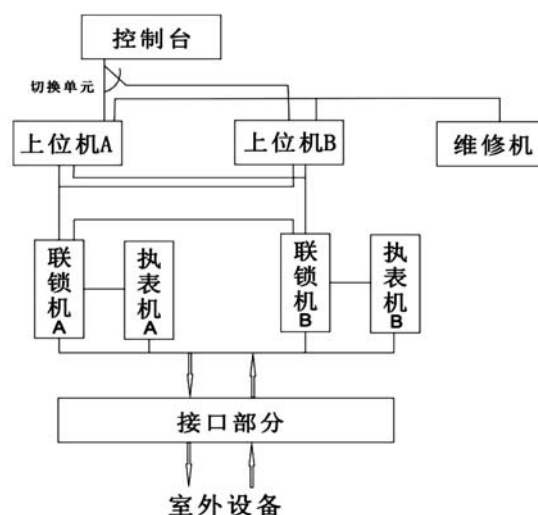


图 1 TYJL- II 型计算机联锁系统结构图

3 TYJL- II 型计算机联锁系统组成

TYJL- II 型计算机联锁系统的组成主要包括监视控制部分和主控部分。监控系统包括控制台和监控机,控制台通过

收稿日期:2023-03-29

基金项目:重庆公共运输职业学院教育教学改革项目“铁路信号工职业资格认证标准与车站自动控制系统维护课程融合的教学改革研究与实践”(项目编号:YSJG202010)。

作者简介:王凤姣(1993-),女,山东青岛人,助教,学士,主要从事铁道信号专业。

中期可逆反应链去除错误率高,后期结果数据不够直观等问题。随着本项目对于燃烧反应数据自动化处理及结果数据可视化方法的逐渐完善,有望被内燃机燃烧反应的相关研究人员广泛使用。

参考文献:

- [1] 贺巧鑫.ReaxFF MD 模拟结果的化学反应网络自动构建及可视化[D].中国科学院大学(中国科学院过程工程研究所),2019.
- [2] 王佳星.基于实验设计的燃烧反应动力学模型不确定性分析研究[D].清华大学,2019.
- [3] 唐钰杰.ReaxFF MD 反应分子体系时空性质分析与可视化[D].中国科学院大学(中国科学院过程工程研究所),2020.

- [4] 蔡东方,陆俞辰,卢东亮,张蕾蕾.碳基固体燃料燃烧反应动力学模型研究进展[J].能源与节能,2021(07):6-9.
- [5] 李晓霞,贺巧鑫,唐钰杰,等.反应分子动力学模拟反应的多层次分析方法 [C]//. 2019 中国化学会第十五届全国计算(机)化学学术会议论文集,2019:115.
- [6] 郭栋梁,王艳芬,李煜,郭瑶瑶,徐锡明,聂俊岚.生物大分子可视化技术综述[J].计算机辅助设计与图形学学报,2021,33(12):1848-1856.
- [7] 张学才,赵春艳.基于动态度的回溯算法求解大值域约束满足问题[J].计算机应用研究,2022,39(05):1427-1431.
- [8] 杨斌,张英佳,李玉阳,陈正,黄佐华,齐飞.面向碳中和的燃烧反应动力学研究进展与展望 [J]. 工程热物理学报,2022,43(08):1993-2008.