

DOI:10.19344/j.cnki.issn1671-5276.2024.06.021

复杂装备全生命周期多源异构数据融合技术研究

曾文驱^a, 王淑莹^b

(西南交通大学 a. 机械工程学院; b. 计算机与人工智能学院, 四川 成都 610031)

摘要:提出一种融合框架和方法,用于解决复杂装备全生命周期数据的多源异构性带来的数据冗余、不一致和索引困难问题。该框架包括自上而下的数据模式融合、自下而上的数据实体模式生成和动态融合全生命周期数据层 3 个步骤。通过建立顶层数据模式、生成数据实体模式并配置主码转换规则和数据对象属性补全规则,实现全生命周期数据层的动态融合。经工程实践验证,该方法能够有效处理复杂装备全生命周期数据的多源异构性,优化数据融合过程。

关键词:复杂装备;全生命周期;融合框架;模式融合;数据融合

中图分类号:TP391 **文献标志码:**A **文章编号:**1671-5276(2024)06-0107-06

Research on Multi-source Heterogeneous Data Fusion Technology for Whole Life Cycle of Complex Equipment

ZENG Wenqu^a, WANG Shuying^b

(a. School of Mechanical Engineering; b. School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: A fusion framework and method are proposed in this study to address the challenges of data redundancy, inconsistency, and indexing difficulties caused by the heterogeneity of multi-source data in the entire lifecycle of complex equipment. The framework includes three steps: top-down data schema fusion, bottom-up data entity schema generation, and dynamic fusion of the entire lifecycle data layer. With the establishment of a top-level data schema, generating data entity schemas and configuring rules for primary key transformation and data object attribute completion, the dynamic fusion of the entire lifecycle data layer is achieved. Engineering practices prove that the proposed method is effective in handling the heterogeneity of multi-source data in the entire lifecycle of complex equipment and optimizing the data fusion process.

Keywords: complex equipment; full life cycle; fusion framework; pattern fusion; data fusion

0 引言

复杂装备是指在重要工业、交通、能源、通信等领域应用的具有复杂结构、高度自动化和智能化的装备,例如高速列车、盾构机、核电装备等。这类装备的特点包括:1)包含大量组件和子系统,需要高度集成;2)结构设计和制造要求复杂;3)研发、制造和维护周期长。

复杂装备涉及众多物理信息系统,导致多维异构信息难以在系统层面上得到统一表达^[1],全生命周期数据关联映射和闭环反馈是复杂装备数字孪生的基石^[2-3]。目前,装备数据融合的研究主要集中在多源异构感知数据融合^[4-5]方法上,包括基于机器学习^[6-7]、深度学习^[8-9]的方法等。这些成果为关键零部件数字孪生研究奠定了基

础,但缺乏对复杂装备结构间关联和履历数据关联的分析。

在全生命周期数据融合研究方面,基于本体的数据集成方法^[10-12]和知识图谱的数据集成框架^[13]已被提出并应用于复杂装备领域。这些方法通过描述全局模式和利用本体知识库来高效地访问多个数据源中的数据。但这些方法缺乏对复杂装备的骨架结构特征和基于产品族配置特征的充分融入。

因此,本文针对复杂装备全生命周期数据的产品结构归属和基于产品族配置设计制造特征,提出复杂装备全生命周期数据集成框架,设计了既有多源异构数据计算的数据实体模式又参照模式自动构建算法,从而实现复杂装备全生命周期数据的一致性融合。

基金项目:国家重点研发计划项目(2020YFB1708000);四川省重大科技专项(2022ZDZX0003)

第一作者简介:曾文驱(1980—),男,广东阳春人,博士研究生,研究方向为智能制造、数据孪生,2464882541@qq.com。

1 复杂装备全生命周期数据模式融合框架

1.1 基于 BOM 的数据关联映射方法

1) 定义 BOM: 描述产品组成关系的数据结构, 通常使用四元组 $B_{OM} = (P, C, R, Q)$ 表示。其中: P 表示产品族或产品; C 表示组件或零部件, 是产品或产品族子部分, 可包含更低层级的子组件或零部件; R 表示组件之间的关联关系; Q 表示组件间的数量关系。

基于 BOM 的复杂装备全生命周期数据关联映射方法: a) 定义 BOM 节点; b) 定义 BOM 节点间关联关系; c) 定义 BOM 节点与实体、各实体间的关联关系。

由于数据的完整性和一致性对复杂装备数字孪生具有重要的意义, 因此采用关系模型, 通过参照完整性来建立实体间的关联约束。

2) 定义参照完整性约束: 设 D_x 和 D_y 代表两类不同对象的数据实体, 其间可建立参照完整性约束:

$$f(D_x, D_y) = \sigma_{D_x.A=D_y.A'}(D_x \times D_y) \quad (1)$$

式中: $f(D_x, D_y)$ 表示 D_x 和 D_y 间的关联映射函数; $D_x \times D_y$ 表示 D_x 和 D_y 两个数据集做笛卡儿积运算; $\sigma_{D_x.A=D_y.A'}$ 表示在 D_x 数据模式的 A 属性和 D_y 数据模式的 A' 属性值相等下的条件选择运算。在上述映射关系中, D_x 称为参照数据关系实体, $D_x.A$ 为 D_x 的参照属性; D_y 称为被参照关系实体, $D_y.A'$ 为 D_y 的主属性集。

1.2 全生命周期数据模式融合框架及实施路径

1) 基于 GBOM 的知识数据模式框架构建

基于产品族的设计制造是一种在多个产品间共享设计和制造资源、知识和经验的方法, 这种方法可以实现产品间的设计和制造的共享及重用, 如图 1 所示。通过建立 GBOM 节点间参照、GBOM 和专业族库、设计规则、维护方法、数据驱动的分析处理模型等可重用的知识类实体间的参照完整性映射, 实现产品族知识模型和数据的关联和一致性约束。

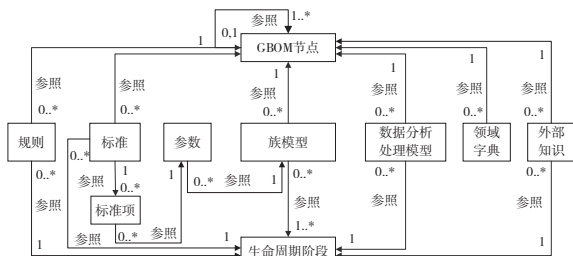


图 1 基于产品族 GBOM 的知识数据模式融合框架

2) 基于 EBOM 的设计数据模式框架构建

设计阶段数据模式融合框架如图 2 所示。在基于产品族的配置设计过程中, EBOM 通常是基于 GBOM 节点选配实例化自动生成, 设计模型和设计参数是在配置设计过程中通过族模型和参数实例化生成, 因此要先建立 EBOM 与 GBOM 节点、模型及参数间参照完整性约束, 其次建立 EBOM 节点与设计阶段零部件模型、二维图册、仿真分析、强度分析等数据的映射。

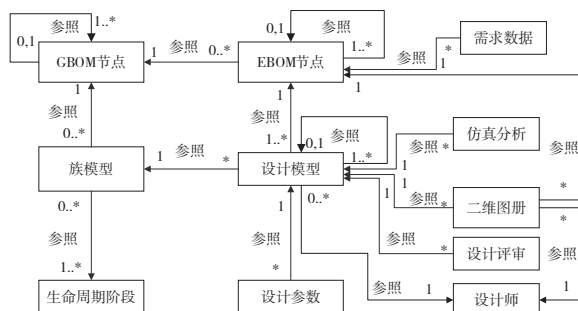


图 2 基于 EBOM 的设计数据模式融合框架

3) 基于 MBOM 的制造数据模式框架构建

制造阶段数据模式融合框架如图 3 所示。对于 EBOM 中的自制件需要设计产品工艺和工序, 每个自制件按工艺排程生产后会按数量形成多个零部件, 基于此可构建 EBOM 节点与 MBOM 节点、工艺和生产任务之间的完整性约束; 其次制造阶段的生产任务、任务排程、生产报工、质检等数据与 EBOM 节点的每个零部件相关, 从而将制造和设计阶段的数据关联。

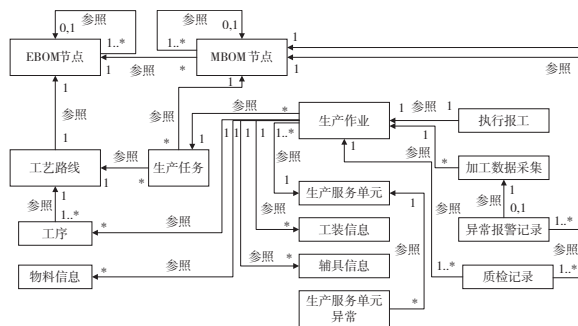


图 3 基于 MBOM 的制造数据模式融合框架

4) 基于 OBOM 的运维数据模式框架构建

OBOM 是针对运维阶段需求, 在 MBOM 的基础上, 结合某些外购件运维需求而构建的, 因此 OBOM 节点基于 MBOM 或 EBOM 实例化生成。基于 OBOM 的运维数据模式融合框架如图 4 所示。建立 OBOM 节点与运维阶段装机档案、运维

记录、感知数据、故障报警、维修环境等数据间的关联映射。

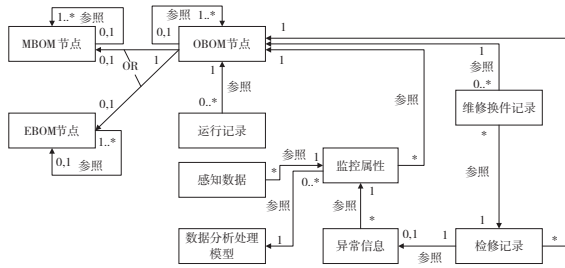


图4 基于OBOM的运维数据模式融合框架

2 基于数据计算的实体数据模式补全

2.1 多源异构数据冗余属性识别与约简方法

不同源系统对同一类数据的同一特征,其属性命名存在差异,按属性数据类型分类设计算法如下。

1) 标称属性约简

设属性 A 和属性 B 是来自不同系统 D_{x1} 和 D_{x2} 的两个属性,属性 A 有 c 个值, $A(a_1, a_2, \dots, a_c)$, 属性 B 有 r 个值, $B(b_1, b_2, \dots, b_r)$, 则属性 A 和属性 B 是否描述同一特征做如下计算:

$$x^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \quad (2)$$

式中: x^2 为卡方检测值; o_{ij} 为观测频度; e_{ij} 为期望频度。

$$e_{ij} = \frac{\text{count}(A=a_i) \times \text{count}(B=b_j)}{n} \quad (3)$$

式中: $\text{count}(A=a_i)$ 表示取 a_i 值的数据样本个数; $\text{count}(B=b_j)$ 为取值为 b_j 的样本个数; n 为数据样本总数。对计算出的 x^2 通过卡方表对比,如果相关,则比较属性 A 和属性 B 的数据类型定义长度,保留数据类型定义长度大的属性,另一个属性标记为冗余。

2) 数值属性约简

设属性 A 和属性 B 是来自不同系统两个属性,通过计算协方差来衡量属性 A 和属性 B 的相关度,即

$$c_{ov}(A, B) = \frac{1}{n} \sum_{i=1}^n (a_i - \bar{A})(b_i - \bar{B}) \quad (4)$$

式中: $a_i \in (a_1, a_2, \dots, a_n)$ 为属性 A 的 n 个样本的值; $b_i \in (b_1, b_2, \dots, b_n)$ 为属性 B 的 n 个样本的值; $\bar{A}$ 和 $\bar{B}$ 分别为属性 A 和属性 B 的均值。

由协方差进一步计算属性 A 和属性 B 的相关

系数:

$$r_{A,B} = \frac{c_{ov}(A, B)}{\sigma_A \sigma_B} \quad (5)$$

计算结果 $r_{A,B}$ 与设定的阈值比对,若超过阈值,则比较属性 A 和属性 B 的数据类型定义长度,保留数据类型定义长度大的属性,另一个属性标记为冗余。

2.2 基于多源异构数据计算的数据实体属性自动补全

对集成框架中的每一个实体,应用数据采集中间件配置其多源异构数据源,提取数据源中数据的属性和数据值,按 2.1 节算法进行冗余属性的约简,用约简后的属性集更新数据实体模式,实现实体属性的自动补全。如图 5 所示,虚线代表右侧数据实体和左侧数据源之间的配置,实线代表通过计算的融合属性集更新实体数据结构,中间部分表示属性计算融合过程。

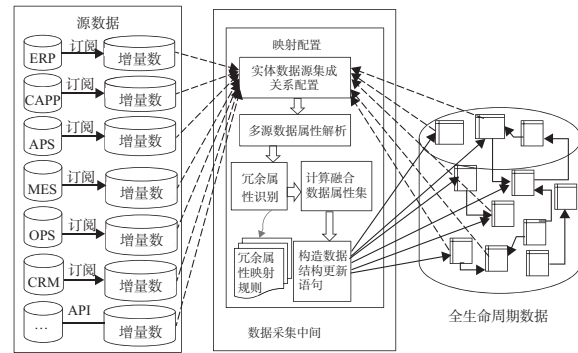


图5 基于多数据源配置的属性融合及实体数据结构更新

多源异构数据属性解析、冗余属性识别及实体数据模式补全算法如下:

输入: 数据实体框架集合

输出: 结构补全后的实体集合

```

1: for 框架中的每一个数据实体 Entityi do
2:   在数据集成中间件中配置多数据源 Sourcei[ ]
3:   if Sourcei[ ]. length = 0 then goto 19
4:   else
5:     提取 Sourcei[ ] 中非主属性及非参照属性集 Attrs[ ]
6:     for 每个属性 Attr do
7:       判断数据类型, 分别加入标称属性队列 NominalList[ ] 和数值属性队列 NumericalList[ ]
8:     end for
9:     for NominalList[ ] 中的每一项 do
10:      与队列中剩余项按(2)带入数据计算, 识别出冗余属性, 从 NominalList[ ] 移除
11:    end for
12:    for NumericalList[ ] 中的每一项 do
13:      与其他项按(5)带入计算, 识别出冗余属性, 从

```

NumericalList[] 移除

14: end for

15: end if

16: 创建 Alter table 语句,用 NominalList[] 和 NumericalList[] 中的属性新建列

17: end for

3 多源异构数据采集及融合存储

在对数据属性进行约简和补全之后,就需要将多源异构数据进行收集和融合,再将对齐后的数据和属性存储入库。该过程如图 6 所示,在图 5 的基础上,基于数据集成中间件进一步配置各实体主码的编码规则,设置源主码与目标实体主码转换规则,然后配置源数据提取、主码转换和数据属性对齐方法;同时将具有参照关联的实体在同一作业中进行编排,以解决实体间存在参照完整性约束的问题。在数据集成中间件中设置任务执行策略,可定期将多源异构数据采集、转换并存储到具有统一模式的数据仓库中,为后续复杂装备数字孪生提供数据支撑。

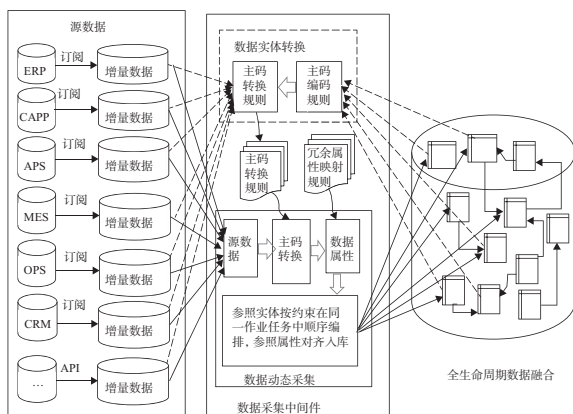


图 6 多源异构数据采集及融合

4 实验验证与分析

以某高速列车产品数据为例,从全生命周期数据融合框架构建、基于多源异构数据解析计算的数据实体属性自动补全和基于 BOM 的全生命周期数据索引 3 个方面进行案例分析以验证方法的有效性。

1) 全生命周期数据融合框架构建

采用 1.2 节介绍的全生命周期数据融合框架构建方法来构建数据实体及参照模式的顶层框架。本实验采用 PowerDesigner16.5 作为框架建模工具,按照图 1—图 4 构建全生命周期数据模式,结果如图 7 所示。

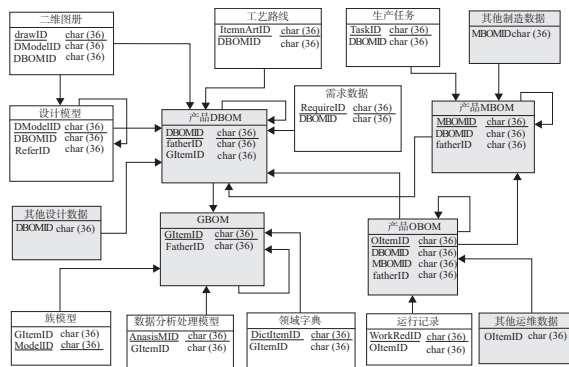


图 7 复杂装备全生命周期数据集成框架实施模型

利用 PowerDesigner 导出工具将框架模型导出为如图 8 所示的数据库 SQL 文件,并在数据库管理系统中执行生成如图 9 所示的实体骨架结构和参照约束。以图中展开的 GBOM 节点为例,其仅包含主码和外码两个属性。

```

/*
/* DBMS name:      Microsoft SQL Server 2012
/* Created on:      2023/2/15 20:44:48
/*
*/

if exists (select 1
from sys.sysreferences r join sys.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('GBOM') and o.name = 'FK_GBOM_REFERENCE_GBOM')
alter table GBOM
drop constraint FK_GBOM_REFERENCE_GBOM
go

if exists (select 1
from sys.sysreferences r join sys.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('二维图册') and o.name = 'FK_二维图册_REFERENCE_设计模型')
alter table 二维图册
drop constraint FK_二维图册_REFERENCE_设计模型
go

if exists (select 1
from sys.sysreferences r join sys.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('二维图册') and o.name = 'FK_二维图册_REFERENCE_产品DBOM')
alter table 二维图册
drop constraint FK_二维图册_REFERENCE_产品DBOM
go

if exists (select 1
from sys.sysreferences r join sys.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('二维图册') and o.name = 'FK_二维图册_REFERENCE_设计师')
alter table 二维图册
drop constraint FK_二维图册_REFERENCE_设计师
go

```

图 8 导出的框架 SQL (部分截取)

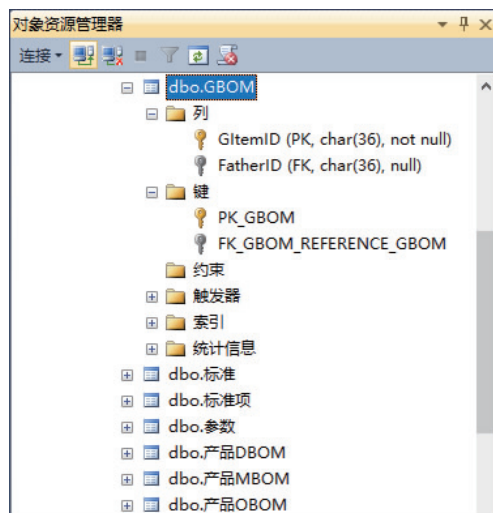


图 9 数据实体参照

2) 基于数据计算的数据实体属性自动补全

对图 7 中的数据实体按照 2.2 节的方法配置要融合的数据源。本实验采用 Kettle 数据集成中间件,配置完成后一次性执行;采集各数据源数据,解析数据属性的数据类型,调用算法 1,生成数据模式更新的 SQL 语句,用融合计算的属性集更新实体数据模式,从而实现补全。

以 GBOM 为例,对基于多源异构数据解析计算的数据实体属性自动补全进行案例分析,所用数据来源于知识库系统和产品设计平台系统。知识库系统 GBOM 模式为:GBOM (id, name, imageaddr, num, introduce, prientid) 包含节点编码、名称、图片、数量、简介和父节点编码 6 个属性,其中非主属性和非参照属性有 { name, image, num, introduce }; 产品设计平台系统 GBOM 模式为 GBOM (code, name), 非主属性和非参照属性有 { name }。将数据输入算法 1,得到 GBOM 属性更新表的 SQL 语句如图 10 所示。在数据库管理系统中执行该 SQL 语句,得到如图 11 所示的更新后的 GBOM。

```
Alter table GBOM
add
name varchar(200),
imageaddr varchar(200),
num int,
introduce text
```

图 10 带入算法 1 计算生成 GBOM 属性更新表 SQL

```
dbo.GBOM
列
  GItemID (PK, char(36), not null)
  FatherID (FK, char(36), null)
  name (varchar(200), null)
  imageaddr (varchar(200), null)
  num (int, null)
  introduce (text, null)
```

图 11 更新后的 GBOM 模式

3) 基于 BOM 的全生命周期数据索引

在完成数据模式融合后,按照第 3 节的方案配置 Kettle 数据采集任务(定时执行),将多源异构数据集成到统一的数据仓库。通过数据中台的可视化界面可查看全生命周期数据。

5 结语

复杂装备全生命周期各个阶段产生的数据无法统一表达,会造成信息孤岛问题,导致资源浪费和数据共享效率低下。为有效解决这些问题,本文提出了复杂装备全生命周期多源异构数据融合框架。该模式的主要贡献在于数据关联映射和数据融合两个方面,研究了面向复杂装备的全生命周期多源异构数据关联映射方法和数据特征统一建模及数据转换方法。

参考文献:

- [1] TAO F, CHENG J F, QI Q L, et al. Digital twin-driven product design, manufacturing and service with big data[J]. The International Journal of Advanced Manufacturing Technology, 2018, 94(9): 3563-3576.
- [2] CHEN L L, ZHANG Y Y, WANG Z P. Logistics service supply chain model applying artificial intelligence and big data analysis [J]. Security and Communication Networks, 2022, 2022: 1575813.
- [3] ZHANG L. Design of a sports culture data fusion system based on a data mining algorithm [J]. Personal and Ubiquitous Computing, 2020, 24(1): 75-86.
- [4] DING W X, JING X Y, YAN Z, et al. A survey on data fusion in Internet of Things: towards secure and privacy-preserving fusion[J]. Information Fusion, 2019, 51(C): 129-144.
- [5] ALAM F, MEHMOOD R, KATIB I, et al. Data fusion and IoT for smart ubiquitous environments: a survey[J]. IEEE Access, 2017, 5: 9533-9554.
- [6] MENG T, JING X Y, YAN Z, et al. A survey on machine learning for data fusion [J]. Information Fusion, 2020, 57(c): 115-129.
- [7] CHEN H P, HU N Q, CHENG Z, et al. A deep convolutional neural network based fusion method of two-direction vibrations signal data for health state identification of planetary gearboxes [J]. Measurement, 2019, 146: 268-278.
- [8] WU J, HU K, CHENG Y W, et al. Data-driven remaining useful life prediction via multiple sensor signals and deep longshort-term memory neural network [J]. ISA Transactions, 2020, 97: 241-250.
- [9] 何逸茹, 刘晨, 杨中国. 面向多源传感器时间序列的时序依赖关联挖掘方法[J]. 小型微型计算机系统, 2021, 42(11): 2307-2311.
- [10] NIANG C, MARKHOFF B B, SAM Y, et al. A semi-automatic approach for global-schema construction in data integration systems [J]. International Journal of Adaptive, Resilient and Autonomic Systems, 2013,

- 4(2):35-53.
- [11] 张春,袁天宁. 针对动车组全生命周期集成管理的多源异构数据融合框架设计[J]. 计算机与现代化, 2017(10):36-41.
- [12] YU Y, SUN L F, WANG S H. Multipart dynamic data integration scheme of industrial chain collaboration platform in mobile computing environment [J]. Wireless Communications and Mobile Computing, 2022, 2022: 1550668.
- [13] KALAYC T E, BRICELJ B, LAH M, et al. A knowledge graph - based data integration framework applied to battery data management [J]. Sustainability, 2021, 13(3):1583.

收稿日期:2023-04-20

(上接第 67 页)

3 结语

本文可以得到以下结论。

1) 盘腔内旋流数沿径向分布趋势基本呈现随半径降低而增大并在最高径向位置处趋于一致的特征。

2) 随着旋转雷诺数的增大,旋流数会减小,并且与调整后的非线性曲线重合部分增大。同时增大的科氏力会抑制流体的震荡,使得整个流场实现更稳定的分层。

3) 随着流量系数的增大,旋流数会增大,并且与自由涡曲线重合部分增大。

4) 罗斯比数的增大会导致气流入射角的减小,而入射角的减小会导致壁面引导入流部分的减少,惯性入流的部分变大,进而减小入口旋流数。

5) 得到了间隙比为 0.263 时, K_i 和 C_{in} 的经验拟合关系式。

6) 本文的模型与实验的误差为 1.915%, OWEN 文献中的理论模型与实验的误差为 19.248%,大大提升了模型预测的精准性。

参考文献:

- [1] 单晶叶. 核心机之路 第四代大推力军用涡轮风扇发动机发展(下)[J]. 航空档案, 2009(9):56-65.
- [2] WEI S, MAO J K, YAN J X, et al. Experimental study on a hybrid vortex reducer system in reducing the pressure drop in a rotating cavity with radial inflow [J]. Experimental Thermal and Fluid Science, 2020, 110: 109942.
- [3] WEI S, YAN J X, MAO J K, et al. A mathematical model for predicting the pressure drop in a rotating cavity with a tubed vortex reducer [J]. Engineering Applications of Computational Fluid Mechanics, 2019, 13(1):664-682.
- [4] FIROUZIAN M, OWEN J M, PINCOMBE J R, et al. Flow and heat transfer in a rotating cavity with a radial inflow of fluid part 1: the flow structure [J]. International Journal of Heat and Fluid Flow, 1985, 6(4):228-234.
- [5] FIROUZIAN M, OWEN J M, PINCOMBE J R, et al. Flow and heat transfer in a rotating cylindrical cavity with a radial inflow of fluid [J]. International Journal of Heat and Fluid Flow, 1986, 7(1):21-27.
- [6] OWEN J M, PINCOMBE J R. Velocity measurements inside a rotating cylindrical cavity with a radial outflow of fluid [J]. Journal of Fluid Mechanics, 1980, 99:111-127.
- [7] 罗翔,冯军. (2007). 一级涡轮盘腔内流体速度场的研究[J]. 燃气涡轮试验与研究(04), 32-35.
- [8] 吉洪湖, CHEAH S C, IACOVIDES H, 等. 旋转盘腔流场速度与压力的实验研究 [J]. 工程热物理学报, 1997, 18(3):300-305.
- [9] 赵亚滨. 用 PIV 测量法研究燃气轮机叶轮与导叶轮间流体的速度场 [J]. 热能动力工程, 1999, 14(4):28-32, 85.
- [10] YAN J X, MAO J K, WEI S, et al. An investigation on the discharge coefficient of compound orifices in rotating disks [J]. Journal of Engineering for Gas Turbines and Power, 2022, 144(4):041007.

收稿日期:2023-03-24