DOI: 10.3901/JME.2021.15.199

基于文本挖掘与神经网络的复杂产品 装配工时估算方法*

刘子文¹ 刘检华¹ 程 益² 庄存波¹

(1. 北京理工大学机械与车辆学院 北京 100081;

2. 上海宇航系统工程研究所 上海 201109)

摘要: 针对卫星等复杂产品装配工时定额主要依靠人工经验确定导致的准确性低、制定速度慢、管理不规范等问题, 提出一种基于文本挖掘与神经网络模型的复杂产品装配工时估算方法。以卫星为例分析了装配工艺数据的特点, 总结了装配工时的影响因素, 并根据工艺特点划分了工艺类别。借助文本挖掘技术对工艺文本特征进行了提取并分类; 在此基础上构建了工时预测神经网络模型, 实现了面向复杂产品装配的定额工时准确估算。设计开发了复杂产品装配工时定额与管理系统, 并在航天某所上线试运行, 应用效果良好, 验证了所提方法的可行性和实用性。

关键词: 复杂产品; 装配; 工时定额估算; 文本挖掘; 神经网络; 工时管理

中图分类号: TP301

Assembly Time Estimation Method for Complex Products Based on Text Mining and Neural Network

LIU Ziwen¹ LIU Jianhua¹ CHENG Yi² ZHUANG Cunbo¹

(1. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081;

2. Aerospace System Engineering Shanghai, Shanghai 201109)

Abstract: Aiming at the problems such as low accuracy, slow formulation speed and nonstandard management caused by artificial experience, an assembly man hour estimation method for complex products based on text mining and neural network model is proposed. Taking satellite as an example, the characteristics of assembly process data are analyzed, and the influencing factors of assembly man hour are summarized, and the process categories are classified according to the process characteristics. Text mining technology is used to extract and classify the process text features; on this basis, the neural network model of man hour prediction is constructed to realize the accurate estimation of quota man hour for complex product assembly. Finally, an assembly man hour quota and management system for complex products is designed and developed, and the system is put into trial operation in an Aerospace Institute. The application effect is good, and the feasibility and practicability of the proposed method are verified.

Key words: complex product; assembly; man-hour quota estimation; text mining; neural network; man-hour management

0 前言

工时定额是指在制造企业中按照规定的工作流程完成一个产品所花费的时间。工时定额是企业计划排产和人员调度的生产指导, 是管理考核的标准

与生产效率的衡量, 是控制人力成本和产品报价的重要依据。在卫星等复杂产品研制生产过程中, 装配过程的重要性日益凸显^[1], 其中卫星装配阶段的工时为产品研制周期的 40%左右^[2], 装配成本超过完整生产成本的 50%^[3]。当前卫星制造企业主要依赖人工经验进行工时定额估算, 效率较低, 主观影响大, 准确性不足; 同时, 工时管理的信息化程度低, 难以满足当前生产管理精细化、数字化和智能化的需求。因此, 有必要研究一种卫星等复杂产品

* 国家自然科学基金(52005042& 51935003)、国防基础科研(JCKY2016204A502)和北京理工大学青年教师学术启动计划资助项目。20200816 收到初稿, 20210311 收到修改稿

装配工时快速估算的方法,并设计开发工时定额与管理系统。

早期的工时定额方法主要是基于人工经验与历史工时数据。1911年,泰勒等对工人作业动作进行分解,开展了基于预定时间标准法(Predetermined time standard, PTS)的工时定额方法研究^[4];此后,学者们对工时影响因素考虑得更加细致和全面,如生产紧急性、精度要求等^[5]。这些方法推动了企业工时定额制定方法的研究,为生产计划的制定和调度排产提供了重要依据。

当前,国内外的研究学者更倾向于结合现代制造技术研究新的工时计算方法,并取得了许多的成果,主要分为以下三类。

(1) 相似工艺分析法,即根据历史工时数据估算出同种或相似工艺的工时数据。例如李亚杰等^[6]提出了基于制造执行系统(Manufacturing execution system, MES)历史数据的实做工时统计方法,分析了工时影响因素并建立了工时预测神经网络。杨青海等^[7]使用基于实例的推理(Case-based reasoning, CBR)法,通过寻找相似实例零件,引入事物特性表,利用回归算法估算工时。李鹏等^[8]利用成组技术对零件分类编码,应用于工时定额中的查表法,提高了工时定额的效率。

(2) 数学分析方法,即建立数学模型揭示影响因素与工时之间的规律。例如晁海涛等^[9]使用支持矢量机算法,建立了机电产品装配过程的工时定额模型。戴健伟等^[11]结合遗传算法优化了BP神经网络,对化工设备设计人工时进行定量预测^[10]。李奇倚等人基于决策树和模型树模型研究工时预测方法,提高了预测精度。习立洋等^[12]建立 PSO-BP 神经网络估算船体分段中间产品额定工时。

(3) 计算机辅助技术,即借助计算机技术对工时定额管理与估算。史恭波等^[13]提出船舶精细化工时管理模式,实现与设计工时数据的有效集成。石晶玉等^[14]对工时定额计算整体框架和数据库开发设计做了创新性研究;洪湖鹏等^[15]开发了带有工时定额计算能力的 CAPP 系统软件。付诚等^[16]基于 SPDM 系统对船舶设计工时进行研究,建立了全面的工时管理体系。

然而,当前大部分的工时预测模型多针对大批量、少品种的零件加工过程和部件装配过程,其工艺信息相对简单,工时影响因素比较提取容易。而对于卫星等复杂产品装配工时定额预测,其装配却存在着装配关系复杂、装配工艺信息多样、数据量庞大、影响因素繁杂等问题,使得工时定额估算比

较困难。为此,本文提出一种基于文本挖掘和神经网络的复杂产品装配工时定额估算方法。通过分析装配工艺数据特点,对工艺文本进行文本挖掘,实现数据特征提取与工艺分类;在此基础上研究工时影响因素,建立神经网络工时估算模型;最后设计开发复杂产品工时定额与管理系统并在航天某所进行上线试运行,验证了所提方法的可行性和实用性。

1 卫星装配工艺分析

1.1 装配工序卡

卫星装配工艺主要以装配工序卡为信息载体,如图 1 所示。该例装配工序卡包括产品名称、产品代号、装入件明细表、工序号、工种、工序名称及内容、设备工装及工具/辅料等装配工艺相关信息。“装入件明细表”中包含有装入零部件的名称、代号和数量信息;“工序名称及内容”中对装配工序操作内容进行了详细的文字描述,含有大量具体的各装配工序过程数据。

装 配 工 艺 卡			产 品 名 称		产 品 代 号		数 量		
			载 荷 舱		TQJ200-2/0-0(c)				
装入件明细表			工 序 号	工 种	工 序 名 称 及 内 容	设备工装及工具/辅料	实 测 记 录	操 作 者	互 检 者
序 号	名 称 代 号	数 量							
1	TQJ200-2/02-0(c) 载 荷 舱-Z下右 隔板	1	M6	总装	下舱装配	三坐标测量仪			
2	TQJ200-2/03-0(c) 载 荷 舱-Z下左 隔板	1	M6.1		将载荷舱底板安装在载荷舱转接环上,用以下紧固件均布紧固: GB/T70.1-2000螺钉M8×20.24件 (工艺件)	力矩扳手			
3	TQJ200-2/07-0(c) 载 荷 舱中隔板	1	M6.2		用三坐标测量机安装精度并记录: 载 荷 舱底板上表面靠近承力框连接处 均匀取不少于9点采样构成的平面与 基准平台的平行度≤0.2mm,若超差 则垫片调整。				
4	TQJ200-2/05-0(c) 载 荷 舱-Y下隔板	1	M6.3		依次将载荷舱中下隔板、载荷舱-Z下 左隔板、载荷舱-Z下右隔板安装在载 荷舱底板上,用M3自制工艺螺钉固 定各隔板两端和中间3点,暂不拧紧。	GD414硅橡胶			
5	TQJ200-2/06-0(c) 载 荷 舱中下隔板	1	M6.4		依次将载荷舱Y下隔板、载荷舱-Y下 隔板安装在载荷舱底板及长隔板上, 分别用M3自制工艺螺钉及 M6×20 工艺螺钉固定各隔板两端和中间3点, 暂不拧紧。	红色泡沫板 δ=20			
更改标记		更改单号	签名	日期	现场 记录				

图 1 卫星载荷舱结构装配工序卡

卫星等复杂产品装配过程属于典型的离散型装配,具有多品种少批量的特点,多数时候是单件生产,其在装配工艺上与常见的大批量机械产品装配存在一定的差异性。主要体现在以下几个方面。

(1) 以文本数据为主。卫星等复杂产品的装配工艺信息虽然有装配工艺卡、二维工艺图纸和三维模型演示视频等数据载体,但唯有装配工艺卡中的数据具有完整性,满足装配指导的要求。这些数据

形式多以非结构化中文文本数据为主。

(2) 完整性。装配工序卡有着装配工序详细的文本记录,包括装配对象、装配动作、装配要求、装配工装、装配步骤次序等,文本记录中的数据量可满足复现完整装配过程的要求。

(3) 复杂性。卫星等复杂产品自身结构特殊,装配对象繁杂,涉及板类、杆类、梁类、结构组件和各种单机产品;其装配动作、步骤多样,会借助各种工装设备和材料辅助装配并且操作复杂;同时在装配过程中,由于卫星产品装配要求严格且部分装配工序的进行前提需要基于上一道工序完成装配验收,多种工种会交替进行处理装配,如漆工、钳工、热处理、质量检测等。这些因素导致了装配工艺的复杂性。

(4) 专业性。工序卡中的文本记录不同于日常生活中使用的文本词汇,具有一定的专业性,信息提取存在较大的困难,常用的、传统的文本数据提取方法不能完全适用,因此要对其文本记录进行数字化信息处理,需要构建相应适配的专业词库进行调用。

(5) 相似性。装配工艺中占据大量的重复性装配动作,例如仪器板的装配,流程基本为预固定第 n 层 m 象限仪器板→精测第 n 层 m 象限仪器板→调试固定第 n 层 m 象限仪器板(n 为 1~4, m 为 1~4),这些重复性装配动作在文本信息的体现中具有一定的相似性,可用于工艺信息归类处理,以便更好地对装配流程进行管理规划。

1.2 影响因素分析与分类

在卫星等复杂产品的装配过程中,涉及到的装配零部件种类繁多、数量庞大,装配要求严格,因此工时定额的影响因素多样。通过查阅与装配工时预测相关的文献和咨询相关技术人员,并结合有关卫星产品装配的工艺文件,依托装配准备过程、装配工作过程、装配检测过程、涂装装箱过程装配四大过程为主线,从装配对象、装配过程和装配动作中提炼出影响因素,以供不同装配工序工时研究参考。

(1) 装配对象。

从装配对象考虑装配工时影响因素,可从重量、体积、形状三个主要方面进行考虑,如表 1 所示。

表 1 装配对象工时影响因素

工时影响因素	具体划分
重量	小、中、大
体积	小、中、大
形状(简易)	板类、梁类、舱体框类、标准件
形状(复杂)	组件、蜂窝芯、后埋件等

(2) 装配过程和装配动作。

装配动作可按装配四大过程划分,如表 2 所示。

表 2 装配动作工时影响因素

过程	动作	动作子类	工时影响因素
装配准备	搬运	抓取	重量、体积、搬运距离、搬运设备
		领取	
		吊装	
	清洁	擦拭	清洁面积、清洁工具
		清洗	
		吸尘	
装配工作	表面处理	热处理	表面处理面积、处理方法
		打磨	
		修挫	修配量、修配工具
	修配	机加工	
		划线定位	重量、定位精度
		孔定位	
装配检测	夹紧	定位销夹紧	重量、配合面数量、定位精度、夹紧数量、使用夹具
		虎口钳夹紧	
		拉绳固定	
	连接	专用夹具夹紧	紧固件数量、紧固件类型、连接长度
		螺栓不拧紧	
		螺栓涂胶	
涂装装箱	喷涂	螺栓拧紧	焊缝长度、焊条直径
		压入	
		焊接	
	密封	精度检测	检测面积、检测工具
		调试	
		质量检验	
涂装装箱	喷涂	表面热处理	喷涂面积
		喷漆	
		贴胶膜	
	密封	腻子密封	密封件数量、密封件类型
		密封圈	
		密封圈	
涂装装箱	装箱	—	体积、重量、装箱工具、保护要求
		—	

通过对卫星装配过程中的装配对象、装配过程、装配动作特点的分析以及工时影响因素的总结,发现不同装配工序的工时影响因素差别较大。为提高神经网络预测效果,根据航天某所的相关技术人员的建议、工艺文件中的实际情况,将工艺划分为主要结构装配、单机结构装配、其他结构装配、装配准备工艺四个类别。主要结构装配中的装配对象为仪器板、水平板、梁类、立板、隔板、拉杆组件、工艺组件等;单机结构装配中的装配对象为 SADA、

发动机、天线安装台等；其他结构装配中的装配对象为蜂窝芯拼接、埋件、基板安装等；装配准备工艺指舱体场地等的清洁、装配体的吊装搬运、表面处理等。各类别主要影响因素如表 3 所示。

表 3 各类工时影响因素特征

类别	工时影响因素
主要结构装配	精度要求、紧固件数量、装配对象体积、连接动作、是否调试检测
单机结构装配	紧固件数量、装配部件数量、精度要求、装配对象体积
其他结构装配	使用工装、密封方式、夹紧定位
装配准备工艺	动作类型、使用工装、重量、体积

2 基于文本挖掘的工艺快速分类

复杂的神经网络模型的训练需要大量的样本数据来确保解决复杂的装配工时多影响因素的预测问题，对于特定的单一装配工序来说，历史工时数据样本少，工时预测影响因素少且简单；而对于多个装配工序来说，历史工时数据样本较多，但由于多种类型的工艺混杂，导致工时预测影响因素多、复杂且工时预测效果不佳。因此本论文通过对具备相同的工时影响因素的相似工艺聚类，累积工艺数量而提高了每个类别的工时数据样本量；同时降低了每类的工时预测问题的难度和影响因素种类，使得每一类的工时预测神经网络无需海量的样本数据，便可取得不错的预测效果，即采取的是“先分类，再预测”的方法。

大批量小品种产品的工艺文件有着良好的结构化数据成分，便于信息提取和运用数学模型进行分析处理；而卫星产品的装配工艺文件以高度非结构化的文本数据为主，数据量大，专业术语多，信息提取难度大。因此需要对非结构化文本数据进行处理，使其能运用于数学模型和计算机系统中。本文使用自然语言处理技术(Natural Language Processing, NLP)中的文本挖掘技术(Text Mining)。工艺分类流程如图 2 所示。

2.1 文本预处理

文本预处理是对作为数据来源的文本信息进行一定的加工处理，将文本信息特征转化为利于文本挖掘工具处理的中间数据形式。以卫星装配工艺文本为处理对象，主要包括以下步骤。

(1) 工艺分词。

分词指对一段文本序列进行切分，得到一个个单独的词汇的过程。工艺分词的目的是将非结构文本信息进行拆分，得到相应的语料，便于后续处理。

对于中文文本信息的分词处理，本文使用 python 语言环境下加载的第三方库 jieba 工具(可通过命令提示符中“pip install jieba”命令下载安装)，它能实现中文分词、自定义词典添加、停用词去除、词性标注等功能。

图 2 工艺分类流程

(2) 专业词库建立。

jieba 工具分词在对卫星装配工艺文本分词时，对于未在原始库中的专有名词无法准确划分，分词效果较差。因此为了提高文本分词效率与准确率，建立自定义装配工艺专业词库是不可避免的。专业词库条目主要由词汇、权值、词性组成，其中“权值 10 000、15 000 和 20 000”分别表示重要程度低、中和高；“词性 n、v”分别表示名词和动词。如表 4 所示。

表 4 专业词库展示

词汇	权值	词性
仪器板	10 000	n
短梁	10 000	n
后连接框	15 000	n
中间框	10 000	n
MS 防松胶	10 000	n
螺接	20 000	v
热控实施	10 000	v

(3) 停用词去除。

在工艺文本中，存在着大量的标点符号、语气词和连接词。这些词汇出现频率高，却和文本的主要信息没有实际关联，属于冗余信息词汇，如“这

个”、“那个”、“于”、“在”等。这些冗余信息词汇在特征提取前需要被剔除。为增强特征提取效果，根据需求建立了停用词表。

抽取一条工艺文本如图 3 展示分词效果。

图3 文本分词效果

对章节一划分的四个类别工艺文件进行文本预处理，处理后形成分词词库。使用词云(wordcloud)展示，其词云中词的字号与词频呈正相关，即词频越大，词的字号越大，如图 4 所示。

图4 四个类别分词词库的词云图

词云图的优点是可以直观显示基于此文档的关键词。可以明显看出，在主要结构装配中，特征词为“仪器板”、“螺钉”、“立板”、“视图”、“螺栓”等；在单机结构装配中特征词为“安装”、“星体”、“天线安装台”、“推力器”、“侧板”、“顶板”等；在其他结构装配中特征词为“产品”、“固化”、“埋件”、“后埋件”、“蜂窝”等；在装配准备工艺中特征词为“轨道舱”、“工艺”、“起吊”、“结构”、“安装”等。

2.2 文本挖掘

文本特征是体现文本信息特征的关键词，为更加准确地提取文档的特征，本文使用 TF-IDF(Term frequency-inverse document frequency)算法对关键词提取。TF-IDF 是一种统计方法，用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。其计算公式为

$$f_{ij} = \omega_{ij} \times \varphi_{ij} \quad (1)$$

$$\omega_{ij} = \frac{n_{ij}}{k_j} \quad (2)$$

$$\varphi_i = \lg \left(\frac{N}{N_i + 1} \right) \quad (3)$$

(1) f_{ij} 为词 i 在工序 j 中的 TF-IDF 值， f_{ij} 值表征词 i 在工序 j 中的重要程度，即 f_{ij} 值越大，词 i 在工序 j 中越重要。

(2) ω_{ij} 为词 i 在工序 j 中的词频(TF)， n_{ij} 为词 i 在工序 j 中出现的次数， k_j 为工序 j 中的词数。

(3) φ_i 为词 i 在语料库中的逆文档频率(IDF)， N 是语料库中工序总数， N_i 是语料库中包含词 i 的工序数， $N_i + 1$ 为了防止分母为零(即语料库中所有工序均不包含词 i)； φ_i 值表征词 i 在语料库中的不常见程度，即 φ_i 值越大，词 i 在语料库中越不常见。

为构建易于处理的文本特征矩阵，将所有工序中的词进行统计，构建含不重复词、词数为 m 的词典并对每个词进行排序编号。文本特征矩阵为

$$A = \begin{bmatrix} F_1 \\ \vdots \\ F_j \\ \vdots \\ F_N \end{bmatrix} = \begin{bmatrix} f_{11} & \cdots & f_{1l} & \cdots & f_{m1} \\ \vdots & & \vdots & & \vdots \\ f_{1j} & \cdots & f_{ij} & \cdots & f_{mj} \\ \vdots & & \vdots & & \vdots \\ f_{1N} & \cdots & f_{iN} & \cdots & f_{mN} \end{bmatrix} \quad (4)$$

式中， f_{ij} ($1 \leq i \leq m$, $1 \leq j \leq N$) 为分词编号数为 i 的词在工序编号数为 j 的工序中的 TF-IDF 值，当分词编号数为 i 的词不在工序编号数为 j 的工序中时，则 f_{ij} 为 0。词典中词数 m 越小，则该文本特征矩阵的维度越低，则计算性能越高；因此应将 f_{ij} 值较高的词留存，剔除 f_{ij} 值低的词，降低词典 m 值。TF-IDF 值计算后，将形成词典集。

2.3 模型构建与评估

为实现文本分类，需要先建立分类模型。本文使用多项式贝叶斯算法(Multinomial naive

Bayes), 对提取了文本特征后的文本以一定规律进行分类。贝叶斯算法具有功能易实现、适用性强、词汇分类效果好的优点。通过对工序文本进行 TF-IDF 值求取、文本特征矢量构建, 忽略了词在文本中的相对位置信息, 保证了贝叶斯算法所要求的条件独立性。

为训练多项式贝叶斯分类模型, 则需先构建用于训练的词袋模型。词袋模型是一种无序的词汇组合, 这种无序的词汇集合忽略了语法规则, 常用于

NLP 技术中的文本分类、信息检索中。

特征词库可表征该类别所含有分词属性, 其表征能力与其含有的词的数量、特点有关。对主要结构装配、单机结构装配、其他结构装配、装配准备工艺四个类别的工艺文件文本预处理后的分词词库进行 TF-IDF 值计算。为避免特征词数过多, 出现过拟合现象, 每类按 TF-IDF 值排序选取前 80 个特征词构建特征词库词袋模型。计算值如表 5 所示, 展示前 10 个词。

表 5 各类别特征词与特征值(前 10 个)

序号	主要结构装配		单机结构装配		其他结构装配		装配准备工艺	
	特征词	特征值	特征词	特征值	特征词	特征值	特征词	特征值
1	视图	0.43	安装	0.4	后埋件	0.25	轨道舱	0.39
2	仪器板	0.33	天线安装台	0.33	产品	0.23	舱门	0.29
3	象限	0.31	侧板	0.3	蜂窝芯	0.2	安装	0.24
4	立板	0.29	确认	0.22	铺层	0.19	结构	0.22
5	螺钉	0.27	螺钉	0.19	尺寸	0.19	工艺	0.19
6	螺接	0.18	顶板	0.17	面板	0.18	舱体	0.19
7	垫片	0.18	推力器	0.16	记录	0.18	基板	0.17
8	角条	0.18	工艺	0.14	含胶量	0.16	起吊	0.14
9	参照	0.17	星体	0.14	埋件	0.15	连接	0.14
10	隔板	0.16	结构	0.14	蜂窝	0.15	包装箱	0.13

同时从该四类工艺文件中各随机抽取 50 条, 一共选取 200 条工艺文本作为测试文本, 对每个文本分词并基于测试文本进行 TF-IDF 值计算, 提取其相应特征值, 作为模型输入进行测试, 对模型进行评估。

使用混淆矩阵对分类模型的分类效果进行评估。将装配工艺的分类预测正确或错误现象看成一个二分类问题, 假如样本真实类别为正类且被预测为正类, 即为真正类(α); 假如样本真实类别为正类且被预测为负类, 即为假负类($\bar{\alpha}$); 假如样本真实类别为负类且被预测为负类, 即为真负类(β); 假如样本真实类别为负类且被预测为正类, 即为假正类($\bar{\beta}$)。

对于分类模型, 我们希望分类效果越准确越好, 即要求 α 与 β 的数量大, 而 $\bar{\beta}$ 与 $\bar{\alpha}$ 的数量小。数量是混淆矩阵的一级指标, 但为了去除样本数量规模的影响, 使用混淆矩阵二级指标准确率(A)、精确率(P)、召回率(Re)和特异度(S)将数量的结果转化为 0-1 之间的比率, 便于进行标准化的衡量。进一步使用三级指标 F1-Score(F_s), 综合精确率和召回率, 以 0-1 的取值对分类模型的效果衡量, 1 为最好, 0

为最差。

公式如下所示, i 代表类别

$$A = \frac{\sum_i (\alpha_i + \beta_i)}{\sum_i (\alpha_i + \beta_i + \bar{\beta}_i + \bar{\alpha}_i)} \quad (5)$$

$$P(i) = \frac{\alpha_i}{\alpha_i + \bar{\beta}_i} \quad (6)$$

$$Re(i) = \frac{\alpha_i}{\alpha_i + \bar{\alpha}_i} \quad (7)$$

$$S(i) = \frac{\beta_i}{\beta_i + \bar{\beta}_i} \quad (8)$$

$$F_s(i) = \frac{2P(i) \times Re(i)}{P(i) + Re(i)} \quad (9)$$

通过对四大类中工艺文件的不断地筛选, 增添专业词库, 提高分词效果, 提高了特征提取的效果, 因此获得了达到预期准确率 90% 以上的分类模型, 分类结果如表 6 所示。

表 6 分类模型测试评估表

指标	主要结构 装配	其他结构 装配	单机结构 装配	装配准备 工艺
α	49	46	48	48
$\bar{\alpha}$	1	4	2	2
$\bar{\beta}$	4	1	4	0
β	146	149	146	150
P	0.92	0.98	0.92	1
Re	0.98	0.92	0.96	0.96
S	0.97	0.99	0.97	1
F_s	0.95	0.95	0.94	0.98
A		0.95		

3 基于神经网络的装配工时定额计算

3.1 神经网络模型构建

对于大批量产品的搬运、螺接等动作，工时定额预测有着简单的数学规律，可以用线性的数学模型模拟；而对于步骤繁杂的复杂产品装配环节来说，其装配工时影响因素多样，规律不明显，无法线性表示，因此本文使用神经网络数学模型来对装配工时定额进行预测。

神经网络模型是一种启发于生物神经网络的分布式并行信息处理的算法数学模型。神经网络由输入层、隐含层和输出层串联构成，而每个层由一个或多个神经元并联组成。神经元收到输入 $x_1, x_2, \dots, x_n$ ，每个输入对于输出拥有不同的影响程度权值 w_i ，进行加权求和后达到一定阈值 $-b$ 才能激活输出，并且经过激活函数 f 的变换，形成神经元输出信号 a ，即 $a = f(\sum_i w_i \cdot x_i + b)$ 。如图 5 所示。

图 5 神经元

前述四个类别中的主要结构装配是卫星装配过程中占时最长、工作量最集中的装配过程，最具有代表性，因此以此为例进行详细说明；而其他三种工艺类别在论文中不作详细说明，是因为可复现论文中建立主要结构装配神经网络工时预测模型的步骤。

以主要结构装配为例说明，首先从工艺数据库

选择一道工序文本，经历文本预处理，输入到贝叶斯分类模型中，得到分类预测结果，指导工时定额员选择相应的工时预测神经网络模型；同时工序文本经历文本预处理后，剔除了无用的冗余信息，突出关键信息，辅助工时定额员提炼神经网络所需要的数据，按式(10)构建工时影响因素输入矢量并输入至神经网络中，得到预测工时数据。流程如图 6 所示。

图 6 装配工时预测模型流程图

对于神经网络模型的建立，主要有如下步骤：

(1) 确定输入、输出层节点数。

神经网络的输入层节点数是工时影响因素的个数，输出层输出信号是工时数据。主要结构装配中有紧固件数量、精度(低、中、高)、连接动作(暂不拧紧、涂胶固定、拧紧)、装配对象、调整测试装配 9 个影响因素，因此输入层节点数为 9，输出层节点数为 1。如下对影响因素进行表征。

紧固件数量(AsmNum)表征形式用数量值 N_u 表示。

装配精度(AsmPrCs)表征形式用矢量 $[P_1, P_2, P_3]$ 表示，其中 $P_1, P_2, P_3 = 0$ 或 1。由于精度的高低对装配工时的影响不是线性的，所以用 $[1, 0, 0]$ 表示低精度， $[0, 1, 0]$ 表示中精度， $[0, 0, 1]$ 表示高精度。

连接动作(AsmAct)表征形式用矢量 $[A_1, A_2, A_3]$ 如下所示，其中 $A_1, A_2, A_3 = 0$ 或 1，对应暂不拧紧、涂胶固定、拧紧三个连接动作方式，存在该动作方式时值取 1，反之取 0：

装配对象(AsmObj)表征形式用 O 来表示装配

对象的体积、重量大小, 大则取值 1, 反之取 0:

调整测试(AsmAdjust)表征形式用 A_d 来表示装配过程中是否需要进行大幅度的调整测试, 是则取 1, 反之取 0:

将表征形式以矢量模式表示, 如下式所示, 且如图 7 所示输入神经网络。

$$I=[N_u, P_1, P_2, P_3, A_1, A_2, A_3, O, A_d] \quad (10)$$

图 7 工时影响因素矢量输入神经网络

(2) 确定隐含层层数和节点数。

在本文神经网络模型为三层神经网络模型, 隐含层节点数根据经验公式确定, 常见的隐含层节点计算公式如下所述。

1) 基于最小二乘法的经验公式

$$h=(0.43mn+2.54m+0.77n+0.35+0.12n^2)^{\frac{1}{2}}+0.5 \quad (11)$$

2) Kolmogorov 定理

$$h=2m+1 \quad (12)$$

3) 适用于大样本量的训练

$$h=\text{int}\left[\frac{n(p-1)}{m+n+1}\right] \quad (13)$$

式中, h 表示隐层节点数, m 、 n 分别是输入层和输出层节点数, p 为样本个数, $\text{int}[]$ 为取整函数。在主要结构装配中, $m=9$, $n=1$, $p=372$ 。因此由三个经验公式得出的隐含层节点数为 6、19、38。

(3) 确定激活函数和初始参数。

对于输出层的激活函数, 使用线性函数; 而隐藏层的激活函数, 选择 relu 函数, 可以有效处理回归问题

$$\text{relu}(x)=\max(x,0) \quad (14)$$

relu 具有收敛速度快, 对回归分析问题适应性好, 克服梯度消失问题等优点, 经过后续训练分析, relu 函数对本文模型有良好适应性。

确定初始参数使用 tensorflow 库中的 random 函数生成 1 号随机数种子、标准差为 1 的随机数权重和偏置矩阵。

(4) 确定损失函数。

损失函数常见的有交叉熵损失(Cross Entropy Loss)、均方误差(Mean Squared Error, MSE)等, 处理回归问题 MSE 有着更大的优势。

$$L=\frac{1}{n}\sum_x\|y-a(x)\|^2 \quad (15)$$

式中, n 为样本总数; x 为样本输入; y 为样本实际值; $a(x)$ 为神经网络对应 x 输入的输出值; L 为损失函数值。为提高模型的泛化能力, 防止过拟合, 对损失函数加入 L2 正则化项, 正则化系数 λ 取 0.01。加入正则化项后如下所示

$$L'=\frac{1}{n}\sum_x\|y-a(x)\|^2+\frac{\lambda}{n}\sum_w w^2 \quad (16)$$

3.2 神经网络优化

为确保工时预测的准确性, 在神经网络结构方面, 寻找最佳隐含层节点数; 在参数优化方面, 寻找合适的优化算法。数据集有 372 条工时数据, 取 90% 作为训练集, 10% 作为测试集。

隐含层节点数有 6、19、38 三种进行比较。优化算法使用随机梯度下降 SGD 算法、动量梯度下降 Momentum 算法、Adam 算法和 RMSProp 算法进行测试比较。SGD 算法学习率 0.0001; Momentum 算法, 学习率 0.0001, 动参 0.05; RMSProp 算法, 学习率 0.1, 动参 0.9; Adam 算法, 学习率 0.001。Batch 值为 8, 为保证模型充分收敛, 训练轮数为 20 000。

对于各算法回归分析的效果, 使用平均绝对误差 $\bar{\delta}$ (Mean absolute error, MAE)、均方根误差 $\bar{\sigma}$ (Root square error, RMSE)、R Squared 三种指标衡量。公式如下所示, 其中 n 为测试数据量, y_i 为真实工时数据, $\hat{y}_i$ 为预测工时数据, $\bar{y}_i$ 为平均真实工时数据。

$$\bar{\delta}=\frac{1}{n}\sum_{i=1}^n|y_i-\hat{y}_i| \quad (17)$$

$$\bar{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2} \quad (18)$$

$$R^2=1-\frac{\sum_{i=1}^n(y_i-\hat{y}_i)^2}{\sum_{i=1}^n(y_i-\bar{y}_i)^2} \quad (19)$$

表 7 列出了神经网络在不同算法与不同隐含层节点数中各评价指标的值。图 8 分别展示了这 12

图 10 实做工时填报和查询界面

下面以仪器板装配、天线安装台装配工艺文件为例展示。首先进行分词操作，仪器板装配工艺文件分词结果如表 8 所示。可以观察到，分词结果并

未出现错误分词现象；若出现，则需要对分词过程中使用的停用词库和专业词库进行修改增添。

表 8 仪器板装配分词展示

工艺内容	分词结果
旋转舱门至III象限对地，领取III象限第一层仪器板	旋转 舱门 象限 领取 象限 第一层 仪器板
使用 14-螺钉 M5x22 和 14-垫片 5、2-螺钉 M6x16 和 2-垫片 6，螺接III象限第一层仪器板与后连接框	14 螺钉 14 垫片 螺钉 垫片 螺接 象限 第一层 仪器板 后连接框
对齐仪器板与后连接框上的象限刻线，暂不拧紧。对标准件进行称重，并记录	对齐 仪器板 后连接框 象限 刻线 暂不拧紧 标准件 称重 记录
⋮	⋮

接着，对仪器板装配、天线安装台装配工艺文件分词后构建分词词库进行 TF-IDF 计算与特征提取，取前 10 个特征词与特征值进行展示，如表 9 所示。最后对实例文件进行分类处理，取实例文件的词袋模型与已构建的分类词袋模型进行匹配，分类结果如表 10 所示。

表 9 实例特征词与特征值(前 10 个)

序号	仪器板装配				天线安装台装配			
	特征词	特征值	特征词	特征值	特征词	特征值	特征词	特征值
1	仪器板	0.6	距离	0.2	天线安装台	0.69	热板	0.15
2	象限	0.26	象限刻线	0.18	工艺	0.38	清洁	0.13
3	端面	0.25	螺钉	0.18	安装	0.17	顶板	0.12
4	视图	0.24	对齐	0.16	星体	0.16	侧板	0.12
5	垫片	0.21	领取	0.16	螺钉	0.16	拆下	0.12

表 10 实例文件分类结果

实例文件	文本数量	预测类别	预测数量	准确度(%)	平均准确度(%)
仪器板装配	117	主要结构装配	114	97.43	95.86
		装配准备工艺	3		
		单机结构装配	25		
天线安装台装配	28	主要结构装配	2	89.28	
		其他结构装配	1		

在实例文件中，仪器板装配的分类结果准确度达到了 97.43%，而因为样本数据过少的原因，天线安装台装配分类结果准确度仅达到了 89.28%。总体来说，分类准确度达到了 95.86%。

现仪器板装配工艺文件分类结果为主要结构装配，

因此使用主要结构装配神经网络对其进行工时预测。首先对仪器板装配工艺文件进行工时影响因素数据提取，将紧固件数量、精度(低、中、高)、连接动作(暂不拧紧、涂胶固定、拧紧)、装配对象、调整测试装配 9 个影响因素按矢量表示。如表 11 中所示。

表 11 工时影响因素提取

序号	工艺内容	矢量表示
1	使用 18-螺钉 M5x22 和 18-垫片 5、2-螺钉 M6x16 和 2-垫片 6，螺栓Ⅳ象限第一层仪器板与后连接框。对齐仪器板与后连接框上的象限刻线，暂不拧紧。	(20,0,1,0,1,0,0,1,0)
2	使用 17-螺钉 M5x16，17-垫片 5 螺栓Ⅲ象限第三层仪器板，对齐仪器板与中间框 2 上的象限刻线，暂不拧紧。	(17,0,1,0,1,0,0,1,0)
3	使用 8-螺钉 M5x12、8-垫圈 5 将 XXX 螺栓在Ⅱ象限第三层仪器板的两端，暂不拧紧。	(8,1,0,0,1,0,0,1,0)
⋮	⋮	⋮

影响因素提取之后，将其作为装配工时定额神经网络模型的输入数据，输出各工序的预测工时，并求得误差，如表 12 所示。在所展示数据中，预测工时相比于实际工时最大误差为 7.33%，最小误差为 0.97%，平均误差为 4.24%，使用神经网络模型预测工时取得了不错的实际应用效果。

表 12 工时预测结果

序号	工时/min	预测工时/min	误差(%)	平均误差(%)
1	83	79.67	4.01	
2	67	62.09	7.33	
3	78	76.51	1.91	
4	38	38.37	0.97	
5	89	82.71	7.07	
6	64	62.09	2.98	4.24
7	51	47.83	6.22	
8	40	38.37	4.08	
9	81	79.67	1.64	
10	64	62.09	2.98	
⋮	⋮	⋮	⋮	

5 结论

(1) 提出了基于文本挖掘的装配工艺分类方法。通过对工艺文本数据进行分词、TF-IDF 值计算、文档特征提取等文本挖掘技术，结合贝叶斯分类算法，构建了工序级工艺分类模型，分类准确率达到 95%。为研究提取工艺信息方法提供了良好的思路。

(2) 提出了基于人工神经网络的卫星装配工时预测方法。对工时数据提取装配工时影响因素，建立了三层神经网络，使用线性函数和 Relu 函数作为激活函数，使用 Adam 算法优化参数，构建了工序级工时预测模型，模型预测指标 R^2 为 0.98，达到了非常好的预测效果。

(3) 设计并开发了复杂产品工时定额和管理系统，并将所提方法的成果体现在具体的软件中，提高了企业对工时定额的估算准确率和管理效率。

参 考 文 献

[1] 刘检华, 孙清超, 程晖, 等. 产品装配技术的研究现状、技术内涵及发展趋势[J]. 机械工程学报, 2018, 54(11): 2-28.

LIU Jianhua, SUN Qingchao, CHENG Hui, et al. The state-of-the-art, connotation and developing trends of the products assembly technology[J]. Journal of Mechanical Engineering, 2018, 54(11): 2-28.

[2] 孙刚, 万毕乐, 刘检华, 等. 基于三维模型的卫星装配工艺设计与应用技术[J]. 计算机集成制造系统, 2011, 17(11): 2343-2350.

SUN Gang, WAN bile, LIU Jianhua, et al. Satellite assembly process design and application technology based on 3D model[J]. Computer Integrated Manufacturing System, 2011, 17 (11): 2343-2350.

[3] FAZIO T, EDSALL A C, GUSTAVSON R E, et al. A prototype of feature-based design for assembly[J]. Journal of Mechanical Design, 1993, 115(4): 369-392.

[4] TAYLOR F W. The principles of scientific management[J]. History of Economic Thought Books, 1911, 7(9): 723-724.

[5] BRUCE G. Which work measurement tool[J]. Manufacturing Engineering, 1999, (3): 18.

[6] 李亚杰, 何卫平, 董蓉, 等. 基于制造执行系统数据采集的工时预测与进化[J]. 计算机集成制造系统, 2013, 19(11): 2810-2818.

LI Yajie, HE Weiping, DONG Rong, et al. Man hour prediction and evolution based on MES data collection[J]. Computer Integrated Manufacturing System, 2013, 19(11): 2810-2818.

[7] 杨青海, 祁国宁, 黄哲人, 等. 基于案例推理和事物特性表的零件工时估算方法[J]. 机械工程学报, 2007(5): 99-105.

YANG Qinghai, QI Guoning, HUANG Zheren, et al. Estimation method of parts working hours based on case-based reasoning and tabular layouts of thing

- characteristics[J]. Acta Mechanical Engineering, 2007(5): 99-105.
- [8] 李鹏, 张树生, 周竞涛, 等. 成组技术在计算机辅助工时定额系统中的应用[J]. 计算机应用, 2003(S1): 42-43.
LI Peng, ZHANG Shusheng, ZHOU Jingtiao, et al. Application of group technology in computer aided man hour quota system[J]. Computer Applications, 2003(S1): 42-43.
- [9] 晁海涛, 徐锋, 许晨辉, 等. 复杂机电产品装配工时定额计算方法研究[J]. 机械制造与自动化, 2019, 48(5): 22-24, 33.
CHAO Haitao, XU Feng, XU Chenhui, et al. Study on calculation method of assembly time quota for complex mechanical and electrical products[J]. Mechanical Manufacturing and Automation, 2019, 48(5): 22-24, 33.
- [10] 戴健伟, 吉华, 杨岗, 等. 基于 GA_BP 算法的化工设备设计人工时预测[J]. 计算机集成制造系统, 2013, 19(7): 1665-1675.
DAI Jianwei, JI Hua, YANG Gang, et al. Based on GA_Artificial hour prediction of chemical equipment design based on BP algorithm[J]. Computer Integrated Manufacturing System, 2013, 19(7): 1665-1675.
- [11] 李奇倚, 王磊. 基于决策树和模型树的作业工时估计方法[J]. 计算机集成制造系统, 2018, 24(2): 309-320.
LI Qiyi, WANG Lei. Estimation method of operation time based on decision tree and model tree[J]. Computer Integrated Manufacturing System, 2018, 24 (2): 309-320.
- [12] 习立洋, 吴娜, 吉永军, 等. 基于 PSO-BP 神经网络的船体分段任务包工时定额模型[J]. 船舶工程, 2020, 42(2): 135-141.
XI Liyang, WU Na, JI Yongjun, et al. Time quota model of ship block task contract based on PSO-BP neural network [J]. Shipbuilding Engineering, 2020, 42(2): 135-141.
- [13] 史恭波, 邵家伟, 张浩, 等. 船舶精细化工时管理模式及应用[J]. 造船技术, 2020(3): 86-92.
SHI Gongbo, SHAO Jiawei, ZHANG Hao, et al. Management mode and application of ship fine chemical industry[J]. Shipbuilding Technology, 2020(3): 86-92.
- [14] 石晶玉, 刘虹, 何苏, 等. 汽车铸件生产工时数理统计计算方法[J]. 吉林工业大学学报, 1998(3): 103-106.
SHI Jingyu, LIU Hong, HE Su, et al. Calculation method of mathematical statistics for production hours of automobile castings[J]. Journal of Jilin University of technology, 1998(3): 103-106.
- [15] 洪湖鹏, 郁鼎文, 张玉峰, 等. 通用化计算机辅助工时系统的研究和开发[J]. 制造技术与机床, 2001(7): 55-58.
HONG hupeng, YU Dingwen, ZHANG Yufeng, et al. Research and development of general computer aided man hour system[J]. Manufacturing Technology and Machine Tool, 2001(7): 55-58.
- [16] 付诚, 邱晓静, 秦肖艇. 基于 SPDM 系统的船舶设计工时管理和应用[J]. 船舶与海洋工程, 2020, 36(2): 65-71.
FU Cheng, QIU Xiaojing, QIN Xiaochuan. Ship design man hour management and application based on SPDM system[J]. Ship and Ocean Engineering, 2020, 36 (2): 65-71.
-
- 作者简介: 刘子文, 男, 1999 年出生。主要研究方向为装配 MES 技术。
mail: liuziwenner@163.com
刘检华, 男, 1977 年出生, 博士, 教授, 博士研究生导师。主要研究方向为产品装配与检测技术。
E-mail: jeffliu@bit.edu.cn
程益, 男, 1994 年出生, 硕士。主要研究方向为智能制造。
E-mail: chengyia1994@163.com
庄存波(通信作者), 男, 1991 年出生, 博士, 博士后。主要研究方向为装配 MES 技术、数字孪生技术。
E-mail: zhuangdavid@bit.edu.cn