

工程知识的交互式获取及融合技术研究

2022 年 6 月

中图分类号：TP182
udc 分类号：621

工程知识的交互式获取及融合技术研究

作者姓名	安鹏
学院名称	机械与车辆学院
指导教师	樊红丽 教授 郝佳 副教授
答辩委员会主席	王武宏
申请学位	工学硕士
学科专业	机械工程
学位授予单位	北京理工大学
论文答辩日期	2022 年 6 月

Research on Interactive Acquisition and Fusion Technology of Engineering Knowledge

Candidate Name:	Peng An
School or Department:	School of Mechanical Engineering
Faculty Mentor:	Prof. Fan
	Associate Prof. Hao
Chair, Thesis Committee:	Prof. Wang
Degree Applied	Master of Science
Major:	Mechanical Engineering
Degree By:	Beijing Institute of Technology
The Date of Defence:	June, 2022

工程知识的交互式获取及融合技术研究

北京理工大学

研究成果声明

本人郑重声明：所提交的学位论文是我本人在指导教师的指导下进行的研究工作获得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京理工大学或其它教育机构的学位或证书所使用过的材料。与我一同工作的合作者对此研究工作所做的任何贡献均已在学位论文中作了明确的说明并表示了谢意。

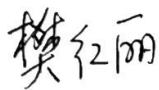
特此申明。

签 名：  日期：2022-6-7

关于学位论文使用权的说明

本人完全了解北京理工大学有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

签 名：  日期：2022-6-7

导师签名：  日期：2022-6-7

摘要

在产品设计领域,代理模型作为一种关键技术愈发受到设计人员关注。然而,由于复杂产品数据的采集成本高、数量少,使得代理模型精度不高。将长期积累的知识融入到代理模型是提高其精度的有效方法之一。然而,工程知识具有类型多、数量多、表示异构等特点,难以统一地融合进代理模型中。为此,本文研究了知识获取、筛选与融合的技术框架及关键技术,支撑知识与数据融合的代理模型构建。具体研究工作包括:

(1) **基于贝塞尔曲线的交互式工程知识获取技术**。该部分针对工程知识缺乏高效的获取方法展开研究,通过分析工程设计案例中工程知识的不同形式,梳理出四种知识类型:属性型、单调型、形状型、公式型。结合贝塞尔曲线等技术构建了交互式知识获取工具,实现工程知识快速获取,为后续知识的检验与融合研究提供知识来源。

(2) **基于多重一致性计算的工程知识检验技术**。该部分针对不同设计专家对工程知识的理解存在认知偏差,会导致工程知识数量多差异大的问题,提出基于梯度信息、基于豪斯多夫距离的两种知识一致性检验方法,并结合 3Sigma 原则完成对异常工程知识的剔除,以避免后续知识的融合中由于异常知识导致的知识信息不准确的问题。

(3) **基于高斯过程的多条目工程知识融合技术**。该部分针对知识筛选后知识数量仍然较多,融入代理模型效率低的问题,提出基于线融合、点融合两种工程知识融合策略,通过高斯过程模型实现融合知识的概率模型表征,在降低工程知识数量的同时有效保留知识信息,以解决多条目工程知识融合代理模型收敛难的问题。

本文结合 benchmark function 和实际应用案例两个方面展开应用验证,结果表明基于贝塞尔曲线的交互式工程知识获取技术能够帮助专家实现工程知识地高效获取;两种知识检验方法分别适用于数据精度高低的不同场景且均能有效剔除异常知识。基于高斯过程的多条目知识融合方法相较于将各知识单独融合进代理模型,能加快模型收敛的同时提升模型的预测精度。

关键词: 代理模型; 工程知识; 贝塞尔曲线; 高斯过程

Abstract

In the field of engineering design, surrogate model technology as a key technology has attracted more and more attention of designers. However, due to the high cost of data acquisition in complex products and the small amount of available data, it is difficult to unify the form of integrating knowledge to build surrogate models. Incorporating long-term accumulated knowledge into a surrogate model is one of the effective ways to improve its accuracy. However, the characteristics of many types of design knowledge, large quantity, and heterogeneity of representation make it still difficult to implement unified knowledge-integrated agent model. This paper puts forward the idea of acquiring and integrating design knowledge for complex equipment, and proposes a technical framework for the entire process from knowledge acquisition to screening and integration for design experts. Specific research work includes:

(1) Interactive Engineering Knowledge Acquisition Technology Based on Bezier Curve. This part conducts research on computable engineering knowledge acquisition methods. By analyzing the different forms of engineering knowledge covered in long-term engineering design cases, it sorts out four levels of knowledge types: attribute type, monotonic type, shape type, and formula type. An interactive engineering knowledge acquisition technique is proposed. According to the characteristics of each type of engineering knowledge, combined with Bezier curve and other technologies to develop knowledge acquisition tools for expert operation, to achieve rapid acquisition of engineering knowledge and provide a knowledge source for subsequent knowledge inspection and fusion research.

(2) Engineering knowledge inspection technology based on multiple consistency calculation. In view of the characteristics of complex product design, due to the large number of experts involved, the differences in cognition, and the large and miscellaneous engineering knowledge, this part proposes two knowledge inspection methods based on data orientation and knowledge orientation, combined with the 3 sigma principle. Complete the screening and elimination of abnormal engineering knowledge, so as to avoid the problem of low surrogate model accuracy due to

inconsistency of potential information between variables embodied in engineering knowledge.

(3) Multi-item engineering knowledge fusion technology based on Gaussian process. This part aims at the problem that after the acquired knowledge is measured and screened by two indicators, although abnormal knowledge will be eliminated, the amount of knowledge cannot be significantly reduced. Two engineering knowledge fusion strategies based on curve fusion and point fusion are proposed, and combined with The Gaussian process model realizes the unified probabilistic model representation of fusion knowledge, which reduces the amount of engineering knowledge and effectively guarantees knowledge information.

In this paper, the application verification is carried out in combination with the benchmark function and practical application cases. The results show that the interactive engineering knowledge acquisition technology based on Bezier curve can help experts achieve rapid acquisition of engineering knowledge; the two knowledge verification methods are suitable for high and low data accuracy respectively. different scenarios, and can effectively eliminate abnormal knowledge. The multi-item knowledge fusion method based on Gaussian process can accelerate the convergence of the model and improve the prediction accuracy of the model compared with the fusion of each knowledge into the surrogate model separately.

Key words: Surrogate model; Design knowledge; Bezier curve; Gaussian process

目录

第 1 章	绪 论.....	1
1.1	研究背景及意义.....	1
1.1.1	研究背景.....	1
1.1.2	研究意义.....	1
1.2	国内外研究现状.....	2
1.2.1	融合知识的代理模型.....	3
1.2.2	工程知识定义与获取.....	6
1.3	课题来源.....	8
1.4	存在问题.....	9
1.5	论文组织结构.....	9
1.6	本章小结.....	11
第 2 章	工程知识的交互式获取及融合技术研究框架.....	12
2.1	引言.....	12
2.2	需求分析.....	12
2.3	技术框架.....	15
2.4	关键技术.....	16
2.4.1	基于贝塞尔曲线的交互式工程知识获取技术.....	16
2.4.2	基于多重一致性计算的工程知识检验技术.....	17
2.4.3	基于高斯过程的多条目工程知识融合技术.....	17
2.5	本章小结.....	17
第 3 章	基于贝塞尔曲线的交互式工程知识获取技术.....	18
3.1	引言.....	18
3.2	基于文本匹配的知识模板构建.....	18
3.2.1	基础知识模板.....	19
3.2.2	基于文本匹配的知识模板构建方法.....	23
3.3	交互式多专家协同的知识获取方法.....	25
3.3.1	基于贝塞尔曲线工程知识修改方法.....	25
3.3.2	基于贝塞尔曲线的单调型与形状型知识获取模板.....	27
3.4	工程知识获取实例.....	29

3.4.1 交互式工程知识获取工具展示	29
3.4.2 工程知识的获取案例.....	30
3.5 本章小结.....	32
第 4 章 基于多重一致性计算的工程知识检验技术.....	33
4.1 引言.....	33
4.2 知识一致性度量指标.....	34
4.2.1 基于梯度信息的工程知识一致性度量指标.....	34
4.2.2 基于豪斯多夫距离的工程知识一致性度量指标	39
4.3 基于 3Sigma 的工程知识筛选方法.....	41
4.3.1 3Sigma 原则	41
4.3.2 知识筛选步骤.....	42
4.4 实验设计.....	42
4.4.1 实验目的.....	42
4.4.2 实验内容.....	42
4.5 实验结果与分析.....	45
4.6 本章小结.....	47
第 5 章 基于高斯过程的多条目工程知识融合技术.....	48
5.1 引言.....	48
5.2 基于遗传算法的曲线信息融合方法.....	48
5.2.1 遗传算法.....	49
5.2.2 基于遗传算法的曲线信息融合方法.....	50
5.3 基于有限点集费马点的点信息融合方法.....	53
5.3.1 有限点集费马点定义.....	53
5.3.2 基于有限点集费马点的点信息融合方法.....	53
5.4 基于高斯过程的多条目工程知识融合.....	55
5.5 实验设计.....	55
5.5.1 实验目的.....	55
5.5.2 实验内容.....	56
5.6 实验结果与分析.....	57
5.7 本章小结.....	58

结论.....	60
总结.....	60
创新点.....	61
展望.....	61
参考文献.....	62
攻读学位期间发表论文与研究成果清单.....	68
致谢.....	69

第 1 章 绪 论

1.1 研究背景及意义

1.1.1 研究背景

工程领域中的复杂产品设计优化问题，存在设计变量多、物理机理复杂、输入输出关系未知等特点^[1,2]。这些特性给设计人员带来了巨大的困难。随着人们设计手段的提高，计算机仿真开始应用各式各样的复杂产品^[3]。但是由于产品结构复杂、设计变量繁多，计算机仿真时需要耗费大量时间，单次计算可能就需数小时、数天甚至数月才能完成。而在产品早期设计阶段，设计方案需要进行快速验证、反复迭代，计算量大、耗时长计算机仿真并不适用^[4-6]。

目前，解决上述问题的有效方法之一是构建代理模型来代替计算机仿真^[7]。相对于计算机仿真，代理模型在保证与原模型近似精度的同时，极大地减小了计算量^[2,7-9]。然而，由于计算机仿真的耗时性、昂贵性，复杂装备设计领域存在可用训练数据少，数据信息覆盖不全面的问题，这使得构建的代理模型的精度难以保证^[10]。考虑到在工程领域，设计专家积累了一定的工程知识，对于设计变量和产品性能之间的变化关系有一定的认识，因此利用工程知识辅助代理模型构建是目前的一个重要方向^[11-13]。

现有的知识获取方法主要从知识工程师角度出发，通过访谈、专家系统等方式来获取知识，在故障检测信息获取等领域能取得很好的效果。然而复杂装备设计过程复杂，涉及的工程知识具有“多且杂”的特点：“多”表现在设计专家人数多，可获取的知识数量大；“杂”表现在专家想法不同，同样的知识会存在认知偏差。因此，如何将差异较大的知识进行检验剔除显得尤为重要。针对以上问题，本文拟梳理复杂装备设计领域工程知识的特点，研究工程知识的交互式获取及检验融合技术，并面向复杂工程设计问题为设计专家提供一种操作简便的导航式工程知识获取工具。

1.1.2 研究意义

在工程设计领域中，工程知识具有类型多、异构且数量众多的特点。工程知识不仅仅反映变量的基本属性，还包括与性能目标之间的准确的公式关系和抽象

的曲线约束关系。专家的主观认知差异会导致知识之间存在偏差,这使得检验知识之间的一致性极为重要。最后针对数量众多的工程知识,在有效降低知识数量的同时保证知识的有效信息也有助于提高代理模型的计算效率。因此,本文针对复杂工程知识的获取、检验与融合展开研究。本文研究内容对融合工程知识构建高精度代理模型具有重要意义,主要包括以下几点:

A. 基于贝塞尔曲线的交互式工程知识获取技术

针对复杂且种类繁多的工程知识,围绕知识的特点实现统一地获取极其重要。本文对设计专家积累的工程知识进行分类定义,通过基于文本匹配的方法构建知识的基本模板,紧接着针对不同类型的工程知识,结合知识模板提出贝塞尔曲线等知识修改方法,实现面向多设计专家协同的工程知识快速获取。

B. 基于多重一致性计算的工程知识检验技术

工程知识包含着设计者的工作实践经验,仅是针对同一设计问题,由于认知差异,专家对于知识的定义也存在不同。复杂装备的设计过程中,由于参与的设计专家人数众多,可获取的工程知识之间差异更加显著。若将其直接融合进代理模型中,会严重影响模型的精度。因此需要将知识做一致性的度量与筛选,从而剔除差异性较大的工程知识以保证知识体现的变量关系趋于一致。本文针对获取的多条工程知识,结合不同的场景提出知识与数据、知识与知识的度量筛选方法:前者是将同时融合进代理模型的数据样本的梯度信息作为参考对知识进行评估筛选,后者则是以知识之间整体相似水平作为筛选依据,实现知识的度量与筛选。

C. 基于高斯过程的多条目工程知识融合技术

完成工程知识的度量筛选后,尽管异常的知识会被剔除,但仍无法显著降低知识的数量。工程知识数量多会影响模型的收敛效率。因此,构建可最大程度表征工程知识共识信息的模型,既能实现降低知识的数量还能保证知识的有效信息。本文提出了基于遗传算法、有限点集费马点的工程知识融合方法来挖掘知识的共识信息。进一步利用高斯过程模型,对融合知识进行建模从而实现知识的统一概率模型表征。

1.2 国内外研究现状

本节内容围绕融合工程知识构建代理模型展开分析。首先是融合工程知识的代理模型的研究历程及其在工程设计领域中的应用,紧接着分析现有的知识获取

的相关方法。

1.2.1 融合知识的代理模型

结合实验数据进行拟合处理, Box 和 Wilson 在 1951 年首次提出多项式相应曲面模型, 这也成为了后来代理模型的雏形。在 20 世纪 90 年代, 随着多学科设计优化的兴起, 面对这类高维度非线性的问题, 响应曲面建模局限性开始显现。在 90 年代末期, 研究重点开始从响应面模型转移到替代近似方法, 包括径向基函数(Radial Basis Function, RBF)^[14], 多元自适应回归(Multivariate Adaptive Regression Splines, MARS)^[15]、克里金方法(Kriging)以及支持向量机(Support Vector Machine, SVM)^[16]等。

表 1.1 常见代理模型构建方法优缺点表

代理模型	优点	缺点
逻辑回归	实现简单、计算量小、速度快	容易欠拟合、只能处理二分类问题
K 近邻	可用于非线性分类, 准确度高	计算量大、样本不平衡问题
支持向量机	低泛化误差、计算复杂度较低	对参数和核函数的选择比较敏感
径向基函数	非线性拟合能力较好、鲁棒性较强	对数据样本有一定的要求
决策树	计算量小、可解释性强	容易过拟合
K 均值	快速、简单	需要指定集群的数量

针对各种代理模型的优缺点, 设计人员结合工程设计问题的不同特征, 逐步衍生出众多不同代理模型构建方法。如表 1.1 所示为常见方法的优缺点。由于代理模型具有计算效率高、成本低的优势, 代理模型已经被广泛地使用于一系列复杂装备设计的工程问题中。李焦赞等人应用多目标进化算法和代理模型技术以解决气动稳健优化设计飞行器环境参数存在不确定性的问题^[17]。韩忠华教授针对大型民机机翼的设计问题, 提出了一系列代理优化算法(Surrogate-Based Optimization, SBO), 大幅缩短了设计周期并且削减了 10% 以上的研制成本, 显著提升了设计效率^[18]。龙腾团队基于径向基函数代理模型用于翼型气动隐身的优化问题, 有效提升翼型升阻比的同时极大降低了重点方位角的雷达散射截面^[1]。杨国伟、于梦阁等人利用代理模型开展了高速列车头型的多目标优化设计, 减少了高速列车的气动阻力, 降低了轮重减载率^[19]。在国外, Qian Z 等人用基于代理模型的进化算法进行飞行器飞行动力学设计, 取得不错的优化效果^[20]。Dodwell,

Haftka, R.T. 团队在结合试验设计方法与优化方法基础上, 使用代理模型技术在航空航天结构优化设计问题上展开了大量研究^[21,22]。Leifur Leifsson 团队在近 10 年将多保真度代理模型技术、代理模型优化技术应用于天线系统、气动结构的设计中^[23,24]。

经过总结发现, 已有方法多应用于装备的一般工况。然而复杂装备的设计涵盖的工况更加多样极端, 在这样的极端工况设计过程中, 仿真计算周期和成本将大幅提升, 难以采集足够的信息。同时一般复杂装备涉及的变量维度较高, 因此有限的信息在高维设计空间分布散乱, 映射规律极其复杂, 训练的代理模型泛化性能欠缺。针对该问题, 学者主要由数据、模型两个方面来提高模型精度。在数据方面, 核心主要是提高数据质量, 包括插值(Interpolation)^[25]、噪声注入(Noise injection)^[26]、数据采样(Data sampling)^[27]、迁移学习(Transform Learning)^[28]、虚拟样本生成(Virtual Sample Generation, VSG)^[29]等方法来扩大数据样本。

然而从数据上着手处理的方法并不能解决少量数据样本的核心问题: 数据体现的变量之间的映射信息不全面。为此, 学者尝试从模型层面解决这一问题, 将额外的变量之间的约束信息融合到代理模型的构建中, 以此来弥补数据信息映射不全的缺陷提高模型的泛化能力。不同于数学问题, 设计专家在长期的工程设计任务中对于变量之间的映射关系有着一定的经验积累, 将这种工程知识融合进代理模型来引导模型构建成为了一个有效的解决方法。

在已有的研究文献中, 用于描述这种工程知识的名称包括: 知识(Knowledge)^[30], 先验知识(Prior knowledge)^[31], 约束(Constraints)^[32], 提示(Hints)^[33]等。

表 1.2 知识与数据融合研究现状表

文献	代理模型类型	知识类型	知识数量	来源专家主观认知
Abu, Y. S., 1990 ^[33]	人工神经网络	提示	1	否
Hinton, 1992 ^[34]	人工神经网络	经验	1	否
Towell, 1994 ^[30]	神经网络	知识	1	否
Tversky, 1996 ^[31]	高斯过程	先验知识	1	否
Sill and Abu, 1997 ^[35]	人工神经网络	提示	1	否
Sill, J, 1998 ^[36]	单调神经网络	提示	2	否
Lang, B, 2005 ^[32]	人工神经网络	约束	2	否

Rasmussen.,2006 ^[37]	高斯过程	先验知识	2-4	否
Daniels, 2010 ^[38]	单调神经网络	提示	1	否
Riihimäki., 2010 ^[39]	高斯过程	经验	1	否
Graeve., 2010 ^[40]	高斯过程	经验	1	否
Simsek, M., 2014 ^[41]	人工神经网络	先验知识	1	否
Lenk, 2017 ^[42]	高斯过程	经验	1	否
Gupta, 2018 ^[43]	人工神经网络	知识	1	否
韩红桂, 2020 ^[44]	人工神经网络	经验	2	否
郝佳, 2020 ^[9]	高斯过程	经验	5	否
郝佳, 2021 ^[45]	进化神经网络	经验	5	否

由表可见，高斯过程和神经网络是用于整合知识的两种主要代理模型。已经有很多研究利用神经网络整合知识和数据，实现的主要机制是对神经网络的参数施加约束。Abu-Mostafa 提出了一种使用神经网络从提示中学习的方法，该提示由一组数据表示^[33]。Sill 进一步发展了这项工作，并提出了将单调性信息纳入神经网络的方法^[35,36]。基于这项工作，Daniels 提出了单调神经网络^[38]，Gupta 提出了具有形状知识约束的神经网络^[43]。与这些方法不同，郝佳等人不仅提出了通过进化神经网络^[45]将有限的仿真数据和工程知识进行整合，还将工程知识作为一种导数信息，结合仿真数据利用高斯过程^[9]相结合。Yu 等人创建了一种名为 KBANN-DT 的基于学习的混合方法，该方法结合了基于知识的人工神经网络（KBANN）和 CART 决策树（DT）。KBANN 用于实现客户需求与产品规格之间的映射，该映射整合了领域理论和历史数据库这两种知识资源^[46]。

以上的研究中形成不同种类的知识都整合到各种代理模型的方法，这使得对于不同的知识，必须设计完全不同的结构和归一化术语。因此，如何针对不同类型的工程知识做统一结构的定义也是需要解决的问题。同时上述研究在融合工程知识时，均假设了其绝对的正确性，这源于初始的研究主要集中在经济、数学领域，知识具有极强的理论依据。近些年随着代理模型向工程设计领域扩展，工程知识的来源多是长期设计专家的工作实践，既包括了变量之间准确的公式关系还包含着一些由于工作实践所产生的对于变量之间抽象约束关系的认知，而这部分工程知识的有效性缺乏检验。除此之外，由于工程设计问题的复杂性，参与设计过程的专家人数众多，知识的来源更加广泛，知识数量随着装备设计的愈加复

杂而剧增，如何提取知识的有效信息以降低知识数量，避免代理模型难以收敛的问题也亟待解决。

1.2.2 工程知识定义与获取

融合工程知识构建高精度代理模型中，工程知识本身有着重要的意义，其表征知识变量之间的关系的准确性极大影响着模型精度。因此首先需要明确工程知识的具体含义，其次围绕不同类型的知识需要分析现有的知识获取方法能否有效展开适用。本文后续将围绕这两部分展开现状分析。

➤ 工程知识

相关的文献中有专门对工程知识的定义和描述：Armaghan 将其定义为人在解决一个实际问题之后获取的知识^[47]。Juarez 认为其是具有一定情节性质的，它与文本知识共同帮助人们解决一系列工程问题^[48]。Azadeh 等人则是将工程知识描述为处理复杂情形过程中，由于具有非线性、时序性以及无法准确量化等特点难以用严密的数学描述的一类实用知识^[49]。随着应用的深入，知识的工程知识有了进一步的描述。Barreto 对知识的含义和成因进行深入分析，指出知识是被构建并相关联的情节，包括个体尝试达成的目标以及在该过程中所接触到的刺激和反应^[50]。Liu 等对知识进行定义：“基于个体的多次实践与思考，以工程设计情境为触发，对事件案例进行概率性关联和引申，所形成的与工程设计问题相关的技术诀窍知识”^[51]。Chan 则是在结合知识与案例之间的关系，通过一系列离散的、标准的属性值来描述知识^[52]。Argote 在知识的研究中提出如何正确的理解知识及其发生的情境至关重要，人应该从多维度对知识及其情境进行详细刻画^[53]。

在国内，张锡爱在机械设计过程中提出这类知识需要与专家或者经验丰富的设计师进行交流并且归纳总结得到^[54]。王定钦等结合模具生产过程中的经验性知识，提出以“问题、原因、解决方案”形式表达的模具经验性知识^[55]。袁晓征在此基础上进一步提出，基于“问题、原因、解决方案”来表达的知识具有不确定性，考虑用产生式规则来表示^[56]。

结合国内外研究对于这类知识的定义和描述，可以发现工程知识来自于实际问题过程中专家的知识积累，体现着设计变量之间极强的约束关系。然而在现有的融合知识代理模型构建框架下，实际的工程知识既包括这类专家的认知知识还包括一些准确的设计公式。同时融合模型的知识类型比较单一，在面向多类别的知识研究中对于知识的定义与结构表征尚未统一。

➤ 知识获取

知识获取是指知识从外部知识源到计算机内部的转换过程：将一些问题求解的知识从专家的头脑中和其它知识源中提取出来，并按照一种合适的知识表示方法将它们转移到计算机中^[57]。然而知识本身来源于人的认知描述，本身是抽象的。因此知识的获取过程中需要一定形式的可视化输入，从而便于计算机进行处理和保存。这里将用于获取知识的可视化输入称之为“载体”。载体本身对于知识获取的有效性有着重要的意义。例如：以“属性-值”的这类键值对的知识列表，便于结合数据挖掘算法实现关联知识的获取；若以自然语言为载体，则知识的内容更加丰富，渠道更多，但难以被计算机识别与获取。如表 1.3 所示为知识获取的主要研究特点，主要从知识的载体、知识的表示形式以及知识的获取方法三个方面。

表 1.3 知识获取研究特点

载体	知识表示形式	获取方法
属性-值	“IF-Then”	机器学习/数据挖掘
Castro-Schez,2013 ^[58]	Ruiz,2014	Liu,2014 ^[51]
Liu,2011 ^[59]	Castro-Schez,2013	Ruiz,2014
Azadeh,2010 ^[49]	Feng,2011 ^[60]	Castro-Schez,2013
访谈	文本	Liu,2011
Frozza,2015 ^[61]	Liu,2011	Wang,2010
Cairo,2012 ^[62]	动作序列	Jin,2006
专家系统	Jin,2006 ^[63]	人工提取
韩,2021 ^[64]	自定义模型	Ruiz,2014
Ruiz,2014 ^[65]	马,2010 ^[66]	Cairo,2012
专家列举	地图	粗糙集
Cheah,2011 ^[67]	Ruiz,2014	Feng,2011
半结构化文本	Cheah,2011	模糊理论
Wang,2010 ^[68]	Wang,2010	王,2020 ^[69]

“属性-值”的键值对形式是常用的一种知识载体，Azadeh 利用模糊理论，结合油泵失效模式数据作为特征输入，将其映射到对应的语义概念，进而定义一系列规则^[49]。Feng 等人将粗糙集理论应用于不确定性知识的获取中，并结合 IF-Then 的决策规则进行表征^[60]。Castro 提出以“类-属性-值”表示领域案例，并提取

出用于案例分类的模糊规则^[58]。

自然语言作为另一种重要的载体，主要包括访谈记录等。Cairo 通过引导受访者进行知识口述，由知识工程师利用一套特殊的符号系统对访谈记录中包含的概念、过程、问题、结构、方案等要素进行表示^[62]。Liu 针对工程设计领域的隐性知识提出一套分为两阶段的知识获取方法：首先通过设定情境的工程师交谈，获取以语言形式存在的知识，紧接着利用关键图（Key Graph）算法对表达关联的知识概念集合进行抽取^[51]。Lee 提出基于访谈记录，构建一系列反映关联信息的认知地图（Cognitive Maps）来实现知识的获取与表示^[70]。

结合具体的问题流程，设计专门的知识获取系统可以帮助人们更加高效的获取更加丰富的知识。Juarez 结合领域本体，将知识获取工具与诊疗过程结合，对包含因果、时序关系的诊断知识进行获取并表示为时序行为模型^[48]。Ruiz 在结合关联规则挖掘算法提取知识基础上，采用概念图（Conceptual Graph）对知识来源和获取结果进行直观展示。除此之外，知识的获取还包括专家列举^[67]以及挖掘半结构化文本^[68]等方法。通过专家列举的方式来获取问题决策过程中相关变量之间的约束关系，再结合计算机算法综合多名专家的意见以形成共识。国内对于知识的研究处于快速发展阶段，主要是在国外研究的基础上进行一定的完善。知识获取技术的研究主要围绕系统设备故障检测^[71]、专家系统构建^[72]等方面，如刘兆政^[73]利用流向图来提取导弹故障诊断的知识；马艳^[66]提出基于规则模板的方法，结合分子材料场景模型来获取知识。张英等提出了一种基于知识地图的知识获取方法，并为专家提供了可视化的操作界面^[74]。

面向工程设计领域，工程知识具有“多且杂”的特点：“多”表现在设计专家人数多，这使得知识的来源多；“杂”表现在由于专家的主观认知差异，对于知识的理解存在认知偏差，使得获取的工程知识存在差异。**目前的知识获取方法难以有效地适用工程设计领域**，尽管工程设计领域有着半结构化的设计文本适用于公式这类准确表达关系的获取，然而工程知识中涉及到变量之间抽象约束的曲线关系等难以被定义，因此亟需一种既能有效处理文本信息又能快熟表征曲线关系的知识获取方法。

1.3 课题来源

本文来源于国家重点研发计划“工程知识与数据融合驱动的复杂产品一体化

智能设计方法”(2021YFB1714500)、GF973 科研项目“基于自学习 xxx 气动外形气动预测技术”(2019-JCJQ-ZD-049-02), 依托项目进行学习和研究。项目均从复杂产品外形的设计过程和企业的实际情况出发, 针对融合知识构建代理模型中的工程知识的获取问题, 基于专家协同操作思路, 围绕工程知识的获取与检验展开研究。具体开展三项研究, 包括工程知识的获取技术、工程知识的检验技术和工程知识的融合技术。在这些技术的支撑下, 设计并开发导航式工程知识获取工具, 为融合知识构建代理模型提供支撑。

1.4 存在问题

融合工程知识是应对工程设计问题中构建高精度代理模型的重要方法, 工程知识很大程度上决定了代理模型的预测精度。然而由于工程知识的数量、类型众多且异构等特点, 使得难以按照统一方式进行获取。通过上述国内外工程知识的研究现状分析, 发现存在以下问题:

(1) 缺乏高效的工程知识的获取方法

工程知识极其复杂且种类繁多, 在复杂装备设计领域多为非结构化的字段或是抽象的形状关系。在工程设计的应用上, 已有研究对知识的分类、表征等远未形成一致的理解与表征。同时工程知识具有极强的主观因素, 传统的知识获取方法难以高效地实现知识的获取。

(2) 缺乏多条目工程知识的一致性检验与融合方法

目前代理模型构建研究大都停留在少量知识的融合上, 实际的复杂装备的设计问题中, 由于参与设计过程的专家人数众多, 可获取的知识数量较多。若直接将各条知识融合进代理模型会大大增加目标数量, 降低收敛效果造成计算量繁杂等问题。同时这类工程知识包含着专家强烈主观认知, 已有的研究缺乏有效性的检验方法。因此将知识进行检验、融合以降低知识数量的同时有效保证知识背后映射的变量之间的约束信息是极为重要的。这对于融入正确的知识构建高精度代理模型有着巨大的现实意义。

1.5 论文组织结构

本文以融合工程知识构建高精度代理模型为背景, 围绕着面向设计专家的工程知识开展系统性研究, 构建了面向设计专家的工程知识获取及融合技术框架,

对基于贝塞尔曲线的交互式工程知识获取技术、基于多重一致性计算的工程知识检验技术和基于高斯过程的多条目工程知识融合技术进行研究，并据此构建了系统工具以进行验证。本文共分为六章，其组织结构如图 1.1 所示。

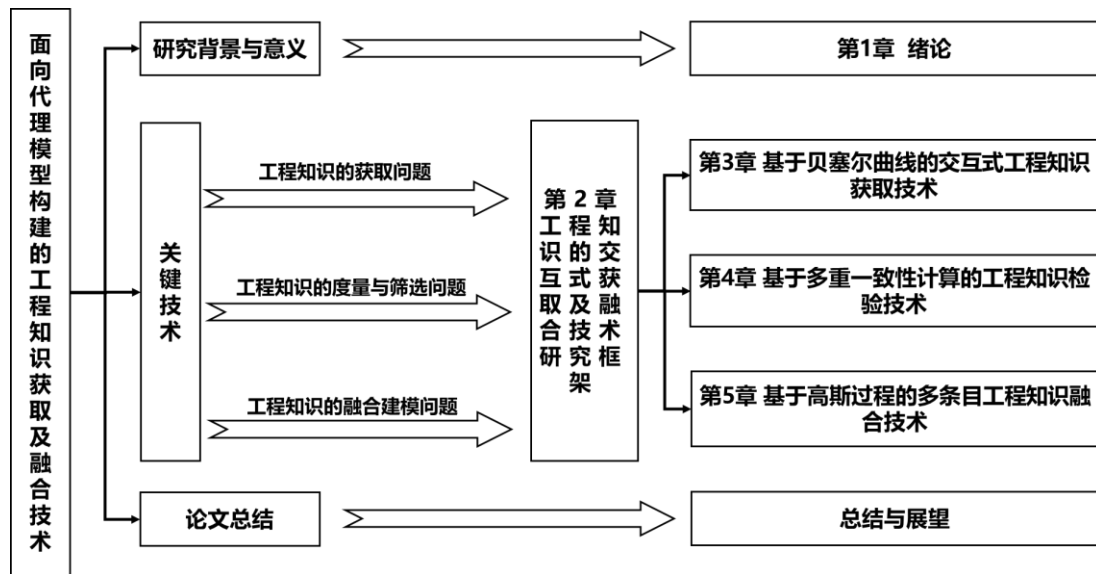


图 1.1 论文组织架构

所示各章节的具体内容安排如下：

第 1 章 绪论。本章介绍了本文的研究背景、研究意义和国内外研究现状以及存在的问题，此基础上提出了论文的主要研究内容、研究技术和组织框架。

第 2 章 工程知识的交互式获取及融合技术研究框架。首先，本章介绍了面向代理模型构建的工程知识获取及融合的需求分析。其次，分析工程知识的基本类型并进行定义。最后，该部分分析了本文研究工作时各研究技术点，介绍了围绕工程知识主要的三个处理环节，阐述了各部分的研究方案及各研究方案之间的关系。

第 3 章 基于贝塞尔曲线的交互式工程知识获取技术。本章针对四种基本的工程知识类型，通过文本匹配构建基本的知识模板，紧接着专家基于贝塞尔曲线等知识修改方法实现知识的协同修改，以完成知识的最终获取。

第 4 章 基于多重一致性计算的工程知识检验技术。本章主要围绕上一章节中获取的知识进行度量，将异常的知识予以剔除。两种方法分别对应知识与知识之间的一致性、知识与数据之间的一致性两种思路，最后结合 3Sigma 原则实现知识的筛选。

第 5 章 基于高斯过程的多条目工程知识融合技术。本章主要针对上一章内

容进行增强,提出利用遗传算法、有限点集费马点思路实现知识的融合,并结合高斯过程模型实现知识统一的概率模型表征。

第6章 总结与展望。主要对本文的研究成果进行了总结,提出了创新点与不足之处以及后续的研究和发展。

1.6 本章小结

本章以工程领域中复杂产品的设计问题为出发点,表明以代理模型去代替仿真计算是一种重要的技术手段。针对复杂产品设计问题,分析了融合工程知识构建高精度代理模型的发展与应用,结合工程知识,总结了国内外围绕知识获取的研究现状,确定了本文的研究思路和研究内容,并提出论文组织结构。

第 2 章 工程知识的交互式获取及融合技术研究框架

2.1 引言

本章组织架构如图 2.1 所示，主要包括以下三部分内容，一是研究内容的需求分析，包括融合工程知识构建代理模型的工程需求、工程知识的分类与定义。二是工程知识交互式获取及检验融合技术的框架。三是本文的关键技术，包括基于贝塞尔曲线的交互式工程知识获取技术（基于文本匹配的知识模板构建方法、贝塞尔曲线）、基于多重一致性计算的工程知识检验技术（豪斯多夫距离度量指标、梯度信息度量指标以及基于 3Sigma 的工程知识筛选方法）、基于高斯过程的多条目工程知识融合技术（遗传算法、有限点集费马点的知识融合策略以及基于高斯过程的融合知识建模方法）。

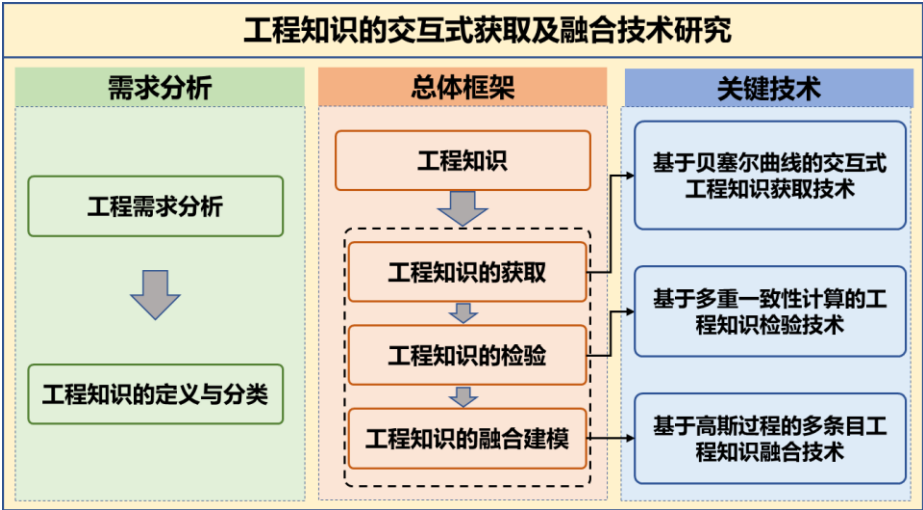


图 2.1 技术框架

2.2 需求分析

在工程设计领域，由于复杂装备的设计涉及到多学科的耦合，设计变量往往存在维度高的问题。设计过程中往往需要大量的仿真计算来实现方案的评估，当涉及到气动特性分析时，由于变量较多、工况复杂，单次的模拟仿真长则花费数月。同时在设计过程中存在反复修改设计方案的问题，设计人员针对设计变量的一次修改就需要重新设计模拟仿真，既延长了时间又增加了计算成本。目前通过数据驱动的代理模型方法来替代计算机仿真是有效应对该问题的方法之一。构建高精度代理模型的目的就是利用算法模型准确学习设计变量与目标性能之间的

映射关系，实现由设计变量的定义到性能目标的快速获取。相比较计算量巨大的仿真实验、代理模型具有计算成本低廉、求解迅速等优势。

然而由于计算机的仿真的耗时、数据采集成本高等问题，复杂产品在设计之初可用于构建代理模型的数据量少。同时复杂产品设计空间巨大，已有的数据仅仅覆盖了少量局部的设计空间，反映出的设计变量与性能目标之间的映射关系不全面。幸运的是在多年的设计过程中，设计人员积累了一定设计经验，对于变量之间的映射规律有一定的认识。因此，将工程知识融入代理模型中，去弥补数据映射信息不全的部分是提高代理模型的预测精度的重要方法。本文围绕面向融合工程知识构建高精度代理模型，针对工程知识这一关键部分进行研究：通过设计面向设计专家的交互操作的知识获取方法，便于获取不同类型的工程知识。在此基础上对获取的知识进行度量检验以剔除异常的知识，最后利用高斯过程模型对知识进行融合建模，实现知识相统一的概率表征。

表 2.1 工程知识类型表

层级	工程知识	解释	示例
1	属性型	描述了设计变量与设计目标之间映射关系的基本属性	输入输出映射的平滑性、不变性、设计目标的范围、设计变量的值域
2	单调型	特定设计变量和设计性能之间的单调关系	如果 x 增加，那 y 会随之增加（或减小）
3	形状型	描述了特定设计变量和设计目标之间的形状关系	如果 x 增加，那 y 会以‘L’型增加或减小（或以‘S’型增加或减小）
4	公式型	描述了基于设计经验或设计公式，多个设计变量和设计目标之间的关系	$y = f(x)$, x 和 y 之间存在具体的数学关系

本文在结合课题组长期围绕复杂装备的工程设计案例实践基础上，将工程知识定义成设计专家依靠多年的设计经验形成的对设计变量和目标性能之间映射关系的认识，并将其分成如表 2.1 所示的四种类型：属性型、单调型、形状型和公式型。

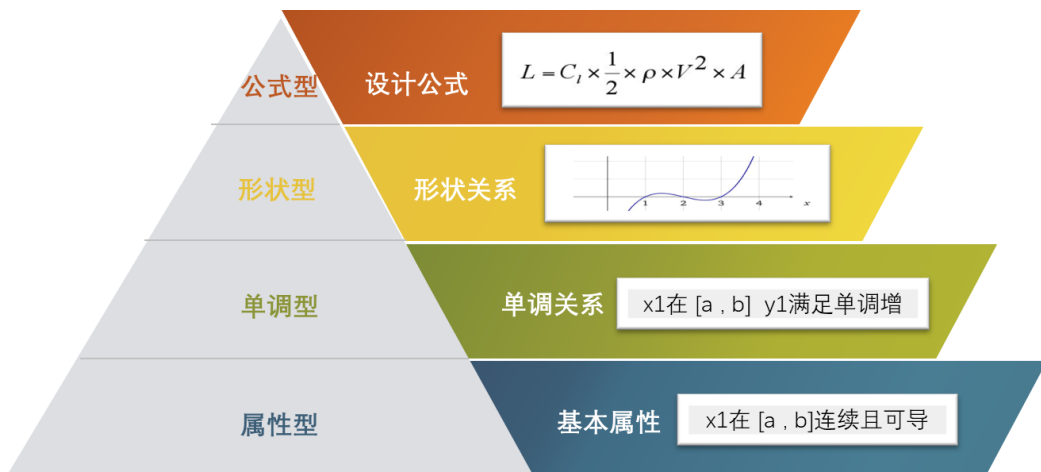


图 2.2 工程知识层次关系

结合各类型知识之间约束关系，本文进一步将以上四种知识划分为图 2.2 所示的 4 个层次。上层知识潜在包含了下层知识核心约束关系，可转换为下层的知识类型：

第一层：基本特性（属性型）

这一层知识主要描述产品设计变量和性能指标，即输入参数与输出参数之间的基本属性，包括参数关系是否连续、定义域、值域范围等信息。这类知识主要用于限制变量的取值范围等。

第二层：单调性关系（单调型）

如图 2.3 所示，这一层知识主要描述产品的设计变量和性能指标，即输入参数与输出参数之间的正负相关性：包括参数间是正相关还是负相关。

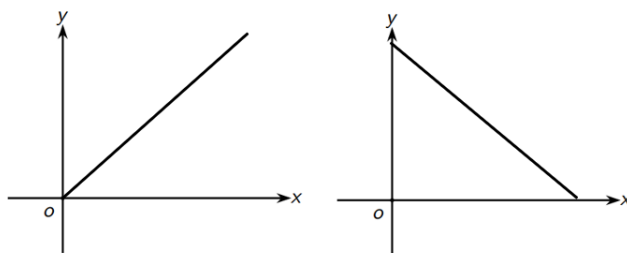


图 2.3 工程知识-单调性关系图

第三层：趋势性关系（形状型）

如图 2.4 所示，这一层知识主要描述产品的设计变量和性能指标，即输入参数与输出参数之间的形状性关系，或者说趋势型关系，主要对某输入参数与某输出参数之间的大致趋势形状，例如 L 型或者双峰型等。这一类的知识同单调类型

的知识为工程案例中常获取的工程知识类型，更多是专家的认知经验，往往带有很大的主观性。这两类知识也作为本文主要研究的知识类型。

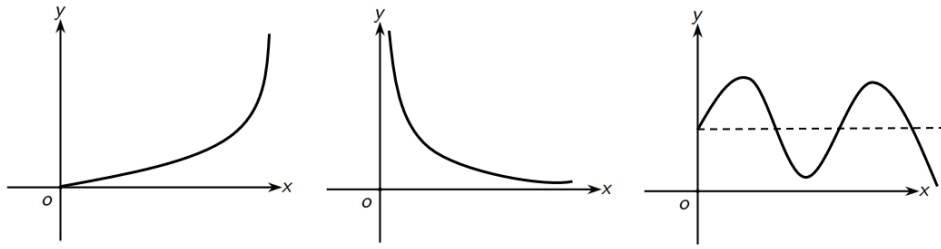


图 2.4 工程知识-趋势性关系图

第四层：参数或部分参数间完整的公式关系（公式型）

这一层知识主要描述产品的设计变量和性能指标，即输入参数与输出参数之间准确的公式关系，例如飞行器横截面积与飞行时受到的阻力具体的函数表达式，由于具有极强的设计理论支撑，一般获取了这类准确的工程知识会直接用于代理模型的构建中。

2.3 技术框架

本文围绕工程知识从分别从三个方面着手研究，以实现知识的获取、检验和融合三个目的。本研究围绕工程知识的总体技术流程如图 2.5 所示。

（1）基于贝塞尔曲线的交互式工程知识获取技术

首先针对工程知识进行分类定义，围绕四种基本的工程知识，提出基于文本匹配的知识模板获取技术。在此基础上，专家利用交互式操作的知识获取技术协同修改模板以实现知识的快速获取。

（2）基于多重一致性计算的工程知识检验技术

这部分目的主要是围绕上一步骤中获取的工程知识，主要面向单调、形状型专家经验的知识类型，提出度量知识一致性的指标。针对其中差异较大的知识进行检验筛选，以避免融入异常知识降低构建代理模型的精度。

（3）基于高斯过程的多条目工程知识融合技术

这部分主要围绕上一步骤中经过筛选的知识，提出知识融合策略来显著降低知识的数量，并结合高斯过程模型对融合的知识进行统一的概率模型表征。

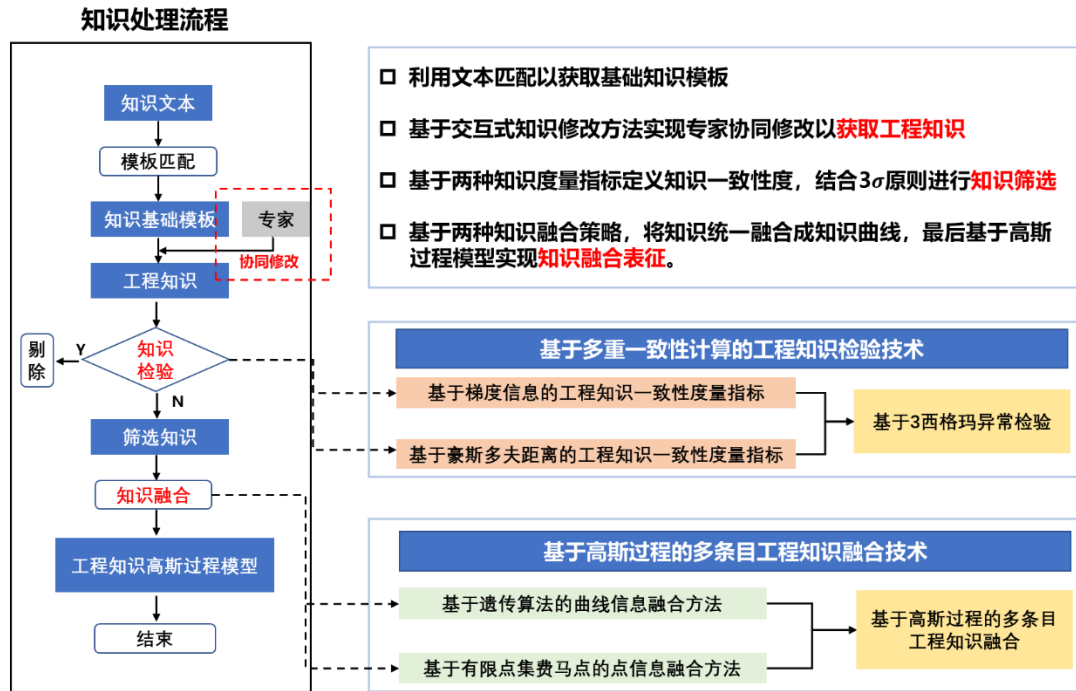


图 2.5 总体技术流程图

2.4 关键技术

在上一节的总体研究框架中，本文提出三项关键技术。围绕着工程知识提出知识获取、检验与融合技术，分别实现知识的获取、知识的筛选以剔除异常知识、知识的融合以降低知识的数量三个目的。

2.4.1 基于贝塞尔曲线的交互式工程知识获取技术

根据本文的定义，工程知识是设计专家在完成工程任务后会积累的设计经验，是对工程问题的主观判断。这些设计经验往往包含了设计参数与目标函数之间的部分映射规律。

首先本文在课题组长期的工程设计案例中，对于工程知识进行整理与分析，提出了属性型、单调型、形状型与公式型四种类型的工程知识。通过分析知识的基本组成结构，首先利用文本匹配构建知识的基础模板，紧接着专家基于交互式操作的知识获取方法完成知识协同的修改。这部分的内容主要为：（1）基于文本匹配的知识模板构建；（2）交互式多专家协同的知识获取方法（基于贝塞尔曲线工程知识修改方法）；（3）工程知识的获取实例。

2.4.2 基于多重一致性计算的工程知识检验技术

融合知识构建高精度代理模型的核心是利用工程知识所表征的变量之间的映射信息去弥补在数据样本覆盖设计空间不全面的缺陷。知识的准确性极大程度影响着最终代理模型的精度。同时面向复杂产品设计过程中,参与的专家人数众多,由于认知的差异,知识具有极高的主观性。因此需要对知识进行度量与检验,将异常的知识予以剔除。本文以知识的度量与筛选为目标,针对不同场景从知识本身的一致性、知识与数据的一致性两个方面进行研究,通过两种度量指标并结合 3Sigma 原则实现对异常知识的检验与剔除。由于变量之间涉及公式类型知识具有准确的表达式且单一,属性型的知识主要表征变量的基本属性不具有可量化的能力,因此检验面向的知识类型为单调、形状型知识。这一部分内容包括:(1) 知识一致性度量指标(基于梯度信息的工程知识一致性度量指标、基于豪斯多夫距离的工程知识一致性度量指标);(2) 基于 3Sigma 的工程知识筛选方法。

2.4.3 基于高斯过程的多条目工程知识融合技术

由于工程设计问题的复杂性,经过筛选后的知识数量仍然较多。现有的融合框架下,知识的数量多会导致模型计算复杂度提升,影响收敛效果。因此,需要最大程度的找到知识之间的共识信息予以表征。本文以最小化融合知识与多知识的距离差异为目标,将知识由曲线形状信息、点坐标信息两种层面进行融合,并通过高斯过程模型对融合知识进行概率建模,实现知识统一的概率表征方法。这部分研究内容为:(1) 基于遗传算法的曲线信息融合方法;(2) 基于有限点集费马点的点信息融合方法;(3) 基于高斯过程的多条目工程知识融合。

2.5 本章小结

本章首先是对研究问题进行需求分析,再以需求分析中的主要研究内容为基础,提出本文的总体技术框架。最后细阐述了技术框架中的三项关键技术,分别是工程知识的获取技术、工程知识的检验技术以及工程知识的融合技术。该三种技术会在 3、4、5 章进行详细介绍。

第 3 章 基于贝塞尔曲线的交互式工程知识获取技术

3.1 引言

结合第二章梳理的四种工程知识的基本特点，如何利用知识模板作为载体完成知识的获取是本章的研究重点。为了实现各类知识的获取，首先针对各类知识的共性部分进行分析，分析发现以上的四类知识核心部分包括针对输入，输出变量本身的信息描述以及变量之间的约束关系的描述两个主要部分。因此，本文围绕知识的获取从一般信息的获取以及约束关系的获取两个方面展开。研究内容如图 3.1 所示：包括基于文本匹配的知识模板构建，通过既定的模板引导专家协同修改处理实现知识的获取。由于各类知识的约束关系并不一致，无法用统一的方法进行获取。因此针对单调、形状型的知识，提出基于贝塞尔曲线的知识获取方法；针对公式、属性的知识类型，提出基于公式编辑工具、文本编辑工具的知识获取方法。

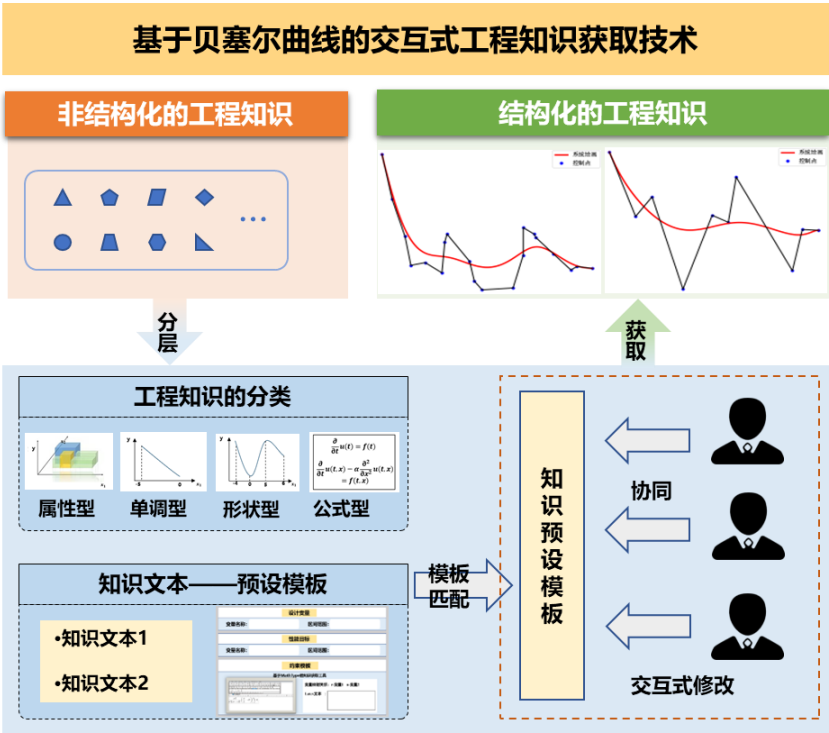


图 3.1 工程知识的获取

3.2 基于文本匹配的知识模板构建

工程设计领域包含大量的设计文档，包含着设计变量与性能目标之间约束关

系的描述。通过对这些文本信息的匹配形成不同的知识基础模板，并将文本重要变量信息提取到模板中能够在帮助设计者快速梳理出知识的基本信息。同时基础的模板能够方便多名设计专家协同修改以实现知识更为高效地获取。因此，本文在针对工程知识进行梳理分类的前提下，提出基于知识文本信息的知识模板匹配方法。首先论述基础的知识模板类型。

3.2.1 基础知识模板

本文通过文本匹配的目的是实现由文本到知识模板的快速转换过程，便于专家利用这些模板快速协同修改以获取知识。因此模板的种类应包含着四种基本的知识类型。本文在此基础上提出相关知识模板作为知识文本匹配依据。首先论述知识模板的基本组成。

➤ 知识模板

结合马艳在知识获取研究中对于规则性知识获取模板的定义，本文提出了基于工程知识特征的知识获取模板。如表 3.1 所示规则性知识获取模板主要包含所属事件、前提条件(即规则前件)、作用目标参数(规则后件作用对象)、目标参数范围(规则后件作用对象取值范围)、作用谓词(规则前件操作)、规则后件(规则性知识中的结果)、返回参数(规则后件返回对象)、返回参数值(规则后件返回对象值)、初始化参数(规则前件有关对象值)^[66]。

表 3.1 规则性知识获取模板

项目	内容	
所属事件	事件内容	
规则前件	规则性知识中的前提条件	
	初始化参数	前提条件的对象
	初始化参数值	前提条件的对象所对应的值
	作用谓词	对初始化参数的操作内容
规则后件	规则性知识中的结果	
	作用目标参数	涉及目标结果的对象
	目标参数范围	作用目标参数的可取值范围
	返回参数	当前前提条件下作用目标参数的返回值

工程知识描述的是设计变量与性能目标之间的映射关系或是变量自身基本属性，基于上述规则模板的基本要素以及工程知识变量关系对其作如下定义，如表 3.2 所示：

表 3.2 工程知识获取模板

项目	内容		工程知识
所属事件	事件内容		所属变量
规则前件	映射关系		设计变量
	初始化参数	设计变量参数名	
	初始化参数值	设计变量参数范围	
	作用谓词	变量描述（属性）	
规则后件	映射关系		性能目标
	作用目标参数	映射变量参数名	
	目标参数范围	映射变量参数范围	
	返回参数	变量映射关系	

基于表 3.2，四种基本知识均可采用“设计变量+设计变量区间+性能目标+性能目标区间+映射关系”进行描述表征。示例如下：

(1) 属性型

变量+区间+描述

如：速度，在[a,b]和[c,d]，满足连续可导

(2) 单调型

设计变量+设计变量区间+性能目标+性能目标区间+单调关系&梯度值

如：速度，在[a,b] 时，温度，在[c,d]，满足单调增，梯度为 0.75

(3) 形状型

设计变量+设计变量区间+性能目标+性能目标区间+形状描述

如：速度，在[a,b] 时，温度，在[c,d]，满足 L 型趋势

(4) 公式型

设计变量+设计变量区间+性能目标+性能目标区间+公式描述

如：速度，在[a,b] 时，温度，在[c,d]，满足 $t=f(v)$

基于定义的知识获取模板，通过文本匹配的核心就是提取变量之间基本的信息，包括变量的名称、取值范围等。围绕着设计变量与性能目标之间的约束关系则是用于区分模板的重要依据。

► 基础约束关系类别

该部分对用于区分各类知识的关键特征词，知识的不同约束类型做出定义，从而基于文本匹配实现由文本信息到基础知识模板的快速获取。单调型的知识描述的是变量之间的正负相关性，本身可以通过直线的梯度值来描述该性质，因此单调型的知识可以归为一种特殊类型的形状型知识。

(1) 属性（详细见 3.4.1 节属性型知识获取系统工具）

关键词：平滑、连续、可导。

约束模板：统一采用可编辑的文本框

(2) 形状（详细见 3.4.1 节形状型知识获取系统工具）

关键词：大概趋势，形状，S，抛物线，指数增长，线性

约束模板：

1) S:

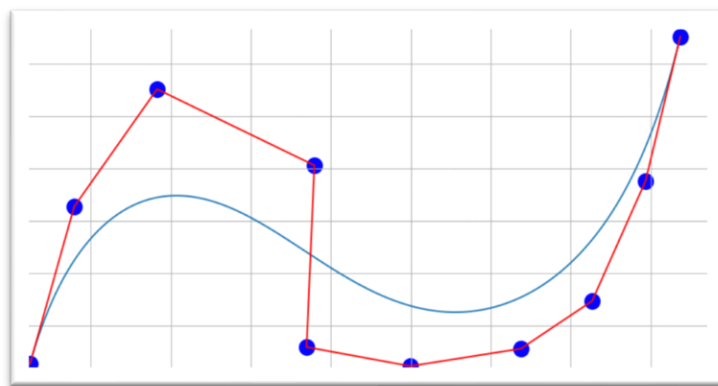


图 3.2 形状-S 型约束模板

2) 抛物线:

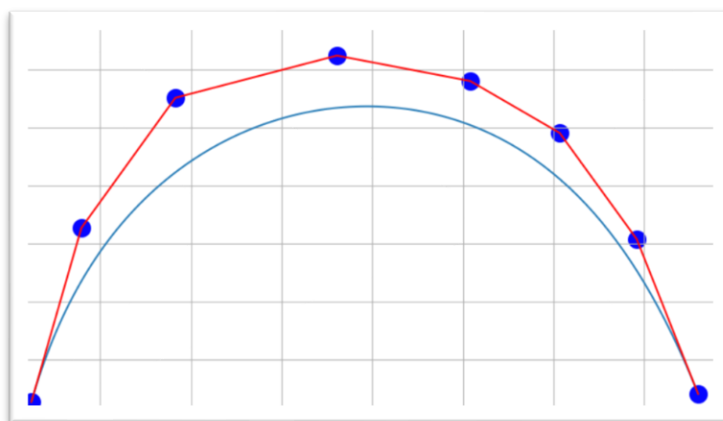


图 3.3 形状-抛物线型约束模板

3) 指数增长:

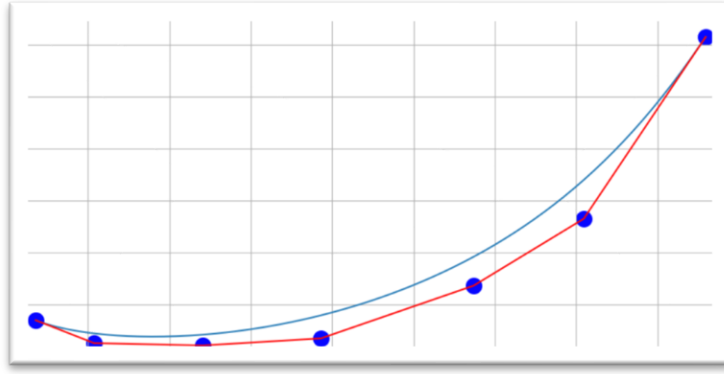


图 3.4 形状-指数型约束模板

4) 线性

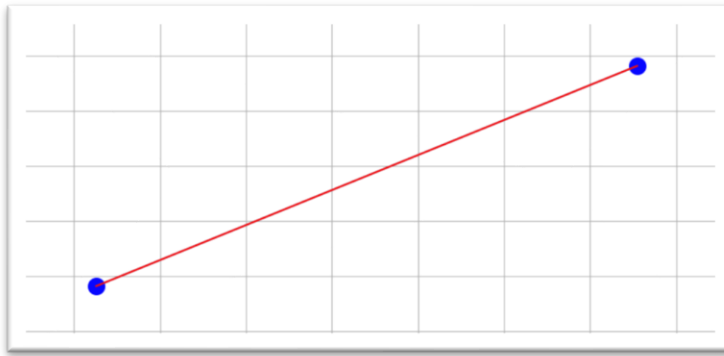


图 3.5 形状-直线型约束模板

(3) 公式 (详细见 3.4.1 节公式型知识获取系统工具)

关键词：公式，计算式，方程，多项式，正余弦，指数，根式，积分，微分

1) 公式-多项式

$$f(x) = k_1 x_1^d + k_2 x_2^m + \cdots k_n x_n^w + h \quad (3.7)$$

2) 公式-正余弦

$$f(x) = A \sin(\omega x + \varphi) + h \quad (3.8)$$

3) 公式-指数

$$f(x) = A e^{-i\omega x} + h \quad (3.9)$$

4) 公式-根式

$$f(x) = \sqrt[n]{Ax^C + B} + h \quad (3.10)$$

5) 公式-积分

$$f(x) = \int_A^B k_1 x_1^d dx + h \quad (3.11)$$

6) 公式-微分

$$f(x) = \frac{\partial kx}{\partial x} + h \quad (3.12)$$

以上所提及的基本约束关系模板中约束关系的表达方法将在 3.3 节内容中进行详细说明, 各类别既在知识的类别上进行区分还针对不同的约束形式进行区分, 便于更高效的知识获取。

3.2.2 基于文本匹配的知识模板构建方法

本方法的核心是通过知识文本文件挖掘变量的名称、取值范围等关键信息, 变量之间描述约束关系的核心关键词等。利用关键词与 3.2.1 中定义的若干约束类型模板相匹配, 从而实现由知识文件到知识模板的构建过程。该方法的核心流程如图 3.6 所示:

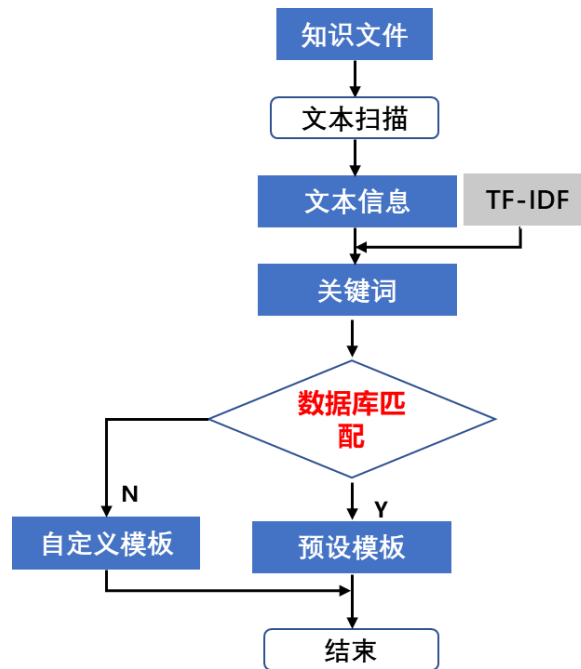


图 3.6 知识模板构建流程

基于文本匹配的知识模板构建方法核心步骤如下:

步骤 1: 文件信息获取

知识文本的信息多以文件的形式进行存储, 文件类型包括 txt、xml、word 等

多种类型的格式。本方法选择 Python 作为文本导入的工具，读取文件的文本信息。

步骤 2：基于 TF-IDF 实现关键词的获取

知识文本中重要的信息包括针对变量自身信息的描述以及变量之间约束关系的描述，因此需要作为关键词提取出。本方法选择 TF-IDF(term frequency-inverse document frequency)来挖掘文本中的关键词，其原理如下：

TF-IDF 算法可拆分为两部分，即 TF 和 IDF。其中，TF (Term Frequency) 是指词频，即词条在文本中出现的频率。为了避免文件的长度对比较产生影响，常对 TF 进行归一化处理，可用公式 3.13 表示：

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (3.13)$$

公式中，分子 n_{ij} 表示词条 t_i 在文件 d_j 中出现的次数，分母 $\sum_k n_{kj}$ 表示所有词条在文件 d_j 中出现的次数和。因此该公式可表示为：

$$TF_{\omega} = \frac{\text{在某一类中词条}\omega\text{出现的次数}}{\text{该类中所有的词条数目}} \quad (3.14)$$

由以上公式可得，一个词条若在文件中出现多次，代表其对文件具有重要作用。但在实际应用过程中，统计出来的 TF 较大的词条常为无实际意义或广泛应用于各类文件的词条，为了从这些词条中选取能够反映当前文件特性的词条，需要引入一个重要性调整系数 IDF (Inverse Document Frequency)，其是指逆向文件频率，是表示词条对不同文件的区分度水平的量。某一词条的 IDF，可以由总文件数目与包含该词条的文件数目相除后取对数得到，可用公式 3.15 表示：

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (3.15)$$

公式中， $|D|$ 表示文件集或语料库中的文件总数， $|\{j: t_i \in d_j\}|$ 表示包含词条 t_i 的文件数目，即 $n_{ij} \neq 0$ 的文件数目，若所选词条不在文件集或语料库中，会导致分母为 0，因此在应用过程中通常使用 $|\{j: t_i \in d_j\}| + 1$ 。因此该公式可表示为：

$$IDF = \log \left(\frac{\text{文件总数}}{\text{包含词条}\omega\text{的文件数} + 1} \right) \quad (3.16)$$

从以上公式可得，IDF 值越大，说明包含某词条的文件数越少，那么该词条可作为文件关键词提取出来。通过 TF-IDF 的方法处理所有的知识文件，提取出

知识文本的关键词。

步骤 3：知识模板的匹配

基于步骤 2 提取的文本关键词，包含着围绕变量自身的描述以及变量之间约束关系的描述。针对 3.2.1 中定义的约束模板的类型进行对应匹配，首先可获取到模板的知识约束类型，若无对应模板匹配则生成自定义模板类型（无约束类型）

本课题组围绕着工程设计等问题，前期已通过 OpenMDAO 可梳理出设计中包含学科、变量等关键信息。在此基础上，本方法针对关键词信息进行对应匹配，提取核心的关键变量。

围绕变量的取值范围的获取，本方法结合关键词利用 python.Re 库中的正则表达式“\d”以及“match”函数挖掘数字信息，实现变量取值范围的获取。由于设计问题中常常设计到数据归一化处理，这里将归一化作为特殊关键词，若匹配上，则设置对应变量取值范围为 0-1。

基于以上三个步骤即实现了由知识文本到知识预设模板的获取，不仅仅包括关键的变量信息还包括取值范围以及对应的变量约束的类型。

3.3 交互式多专家协同的知识获取方法

工程知识的来源的是设计者长期的工作积累的认知，基于文本匹配获取的知识模板并不能准确表达知识。同时代理模型构建人员往往不具有该领域特定的知识背景，并且在沟通和思维方式上与领域专家存在偏差，无法准确的把握知识的内涵，这使得获取到的知识可信度也无法保障。因此本文在 3.2 节中通过文本匹配获取不同类别的知识模板前提下，结合各知识约束关系的核心，提出不同的交互方式针对模板进行修改，以实现面向专家操作的知识快速获取。

结合公式类型知识约束的核心是变量之间数学表达式，属性类型的知识则是变量的取值范围以及是否平滑等文本的描述，相关的技术可选择可编辑的公式编辑器、以及富文本编辑器实现。本文不引申展开两种知识交互的操作修改方法，主要针对形状、单调类型的知识，提出基于贝塞尔曲线的工程知识修改方法。

3.3.1 基于贝塞尔曲线工程知识修改方法

形状类型知识获取的核心是针对曲线的编辑操作，由于贝塞尔曲线能够准确量化描述变量之间特定形状的影响规律，因此本文提出基于贝塞尔曲线的工程知

识的获取方法来实现形状关系的获取。首先论述贝塞尔曲线的基本原理。

➤ 贝塞尔曲线

贝塞尔曲线是应用于二维图形应用程序的数学曲线，其原理依据的是伯恩斯坦多项式（由伯恩斯坦多项式可证明：在 $[a, b]$ 区间上所有的连续函数都可以用多项式来逼近，实现一致收敛），这使得贝塞尔曲线可在指定区间内表征任意的连续函数。现以三阶贝塞尔曲线为例，说明其基本原理：

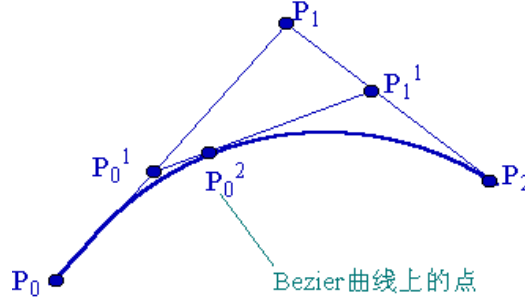


图 3.7 三阶贝塞尔曲线

如图 3.7 所示，设 P_0 、 P_0^2 、 P_2 是一条抛物线上顺序三个不同的点。过 P_0 和 P_2 点的两切线交于 P_1 点，在 P_0^2 点的切线交 P_0P_1 和 P_2P_1 于 P_0^1 和 P_1^1 且有如下比例成立：

$$\frac{P_0P_0^1}{P_0^1P_1} = \frac{P_0^1P_0^2}{P_0^2P_1^1} = \frac{P_1P_1^1}{P_1^1P_2} \quad (3.17)$$

当 P_0 、 P_2 固定，引入参数 t ，令上述比值为 $\frac{t}{(1-t)}$ ，即有：

$$\begin{aligned} P_0^1 &= (1-t)P_0 + tP_1 \\ P_1^1 &= (1-t)P_1 + tP_2 \\ P_0^2 &= (1-t)P_0^1 + tP_1^1 \end{aligned} \quad (3.15)$$

进一步推导：

$$P_0^2 = (1-t)^2P_0 + 2t(1-t)P_1 + t^2P_2 \quad (3.19)$$

当 t 每取一个 $[0,1]$ 的值时，都会有一个点 P_0^2 ，由一系列 P_0^2 构成的拟合曲线即为由控制点 P_0 、 P_1 、 P_2 拟合的贝塞尔曲线。进一步引入多个控制点时，可整理得到关于 n 个控制点的 $(n-1)$ 阶贝塞尔曲线表达式：

$$B_{(t)} = \sum_{i=0}^n C_n^i P_i (1-t)^{n-i} t^i, t \in [0,1] \quad (3.20)$$

其中 $B_{(t)}$ 为 t 取某值下的拟合点， n 为控制点的数量， C_n^i 为排列组合， P_i 为第 i

控制点坐标, t 为取值在 $[0,1]$ 的一系列值。

➤ 基于贝塞尔曲线的约束关系修改方法

结合贝塞尔曲线的表达式可以发现拟合曲线由一系列控制点 P_i 所定义, 曲线上点的数量由 t 所决定。本文利用拟合曲线由控制点直接定义这一特点, 开发了直接由专家操作的交互式知识获取工具。利用知识获取工具, 专家通过增减控制点的数量, 拖动控制点的位置, 经过贝塞尔曲线公式可实时得到拟合曲线。专家完成知识绘制后, 知识获取工具保存控制点坐标, 从而完成知识获取整个过程。

如图 3.8 所示为专家操作的知识形状约束关系获取示例, 当由图 3.8-(a)拖动控制点 2 至图 3.8-(b)位置时, 拟合曲线由图 3.8-(a)改变成 3.8-(b), 依次拖动控制点 4 形成图 3.8-(c)、拖动点 6 形成 3.8-(d)拟合曲线。

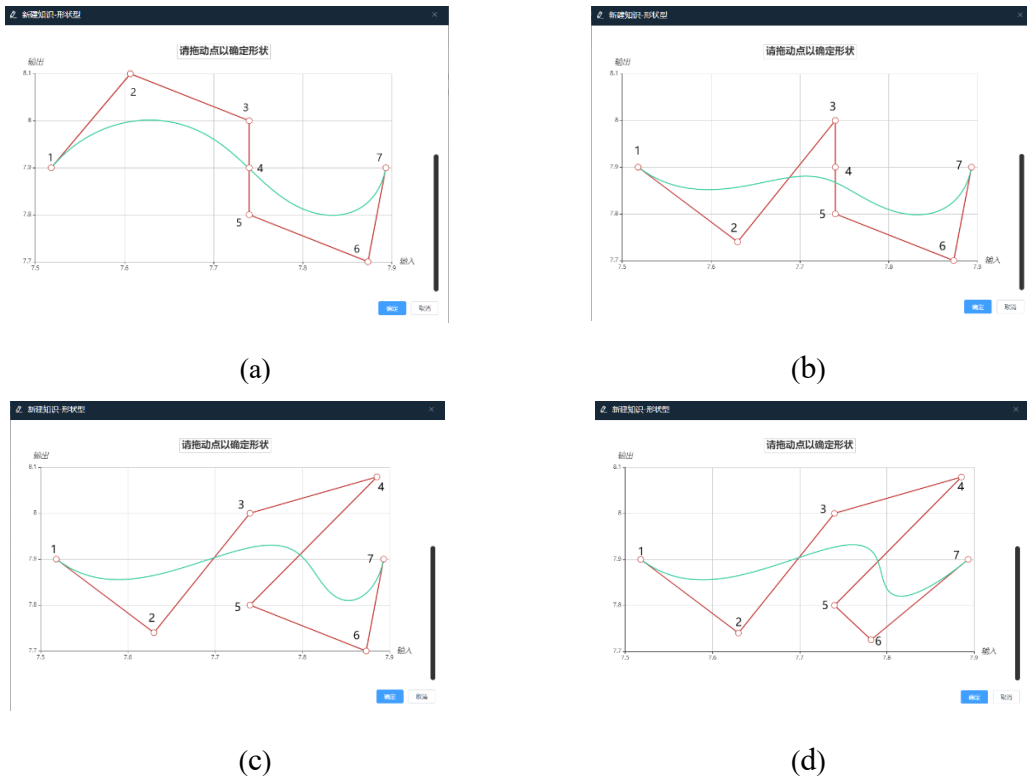


图 3.8 基于贝塞尔曲线的知识获取交互示例

3.3.2 基于贝塞尔曲线的单调型与形状型知识获取模板

结合贝塞尔曲线的约束关系获取方法, 考虑到单调型的知识可以通过直线进行定义表征, 此时贝塞尔拟合曲线的控制点仅有首尾两个如图 3.9 所示, 通过坐标信息计算斜率即可获得该条知识的梯度信息, 结合正负性来表达其单调关系。


设计变量	
变量名称:	区间范围:
性能目标	
变量名称:	区间范围:
约束模板	
基于贝塞尔曲线的知识获取工具	
	控制点数量: 2 梯度值: $\frac{\Delta y}{\Delta x}$

图 3.9 单调型知识获取模板示意图

如图 3.10 所示即为形状型知识获取模板，不同于单调型将控制点的数量限制为两个，此时专家可以通过输入控制点的数量来实现控制点的删减，同时通过点击选择某一控制点进行拖动可实现点位置的改变从而实现拟合曲线的实时定义。以上两个模板在规则模板的内容上补充了针对变量之间约束关系的可操作的工具，实现了这两类工程知识的快速修改。

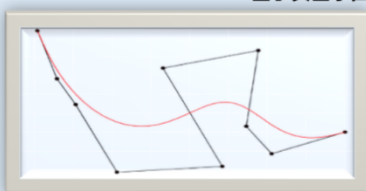
设计变量	
变量名称:	区间范围:
性能目标	
变量名称:	区间范围:
约束模板	
基于贝塞尔曲线的知识获取工具	
	控制点数量: 拟合曲线区间: $x \quad , \quad \quad y \quad , \quad $

图 3.10 形状型知识获取模板示意图

3.4 工程知识获取实例

3.4.1 交互式工程知识获取工具展示

围绕着工程知识的获取，本文开发了相关系统获取工具。本节将主要围绕工程知识的获取部分展示四类知识的系统交互页面。

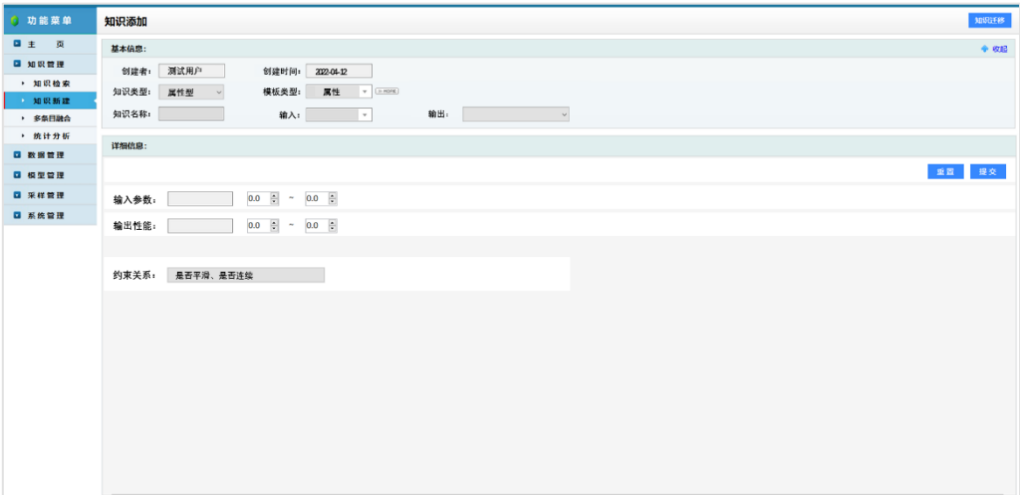


图 3.11 属性型知识新建系统页面

图 3.11 所示的为系统属性型知识的获取页面，主要针对变量名称、取值范围以及是否平滑、连续的基本属性进行修改定义。

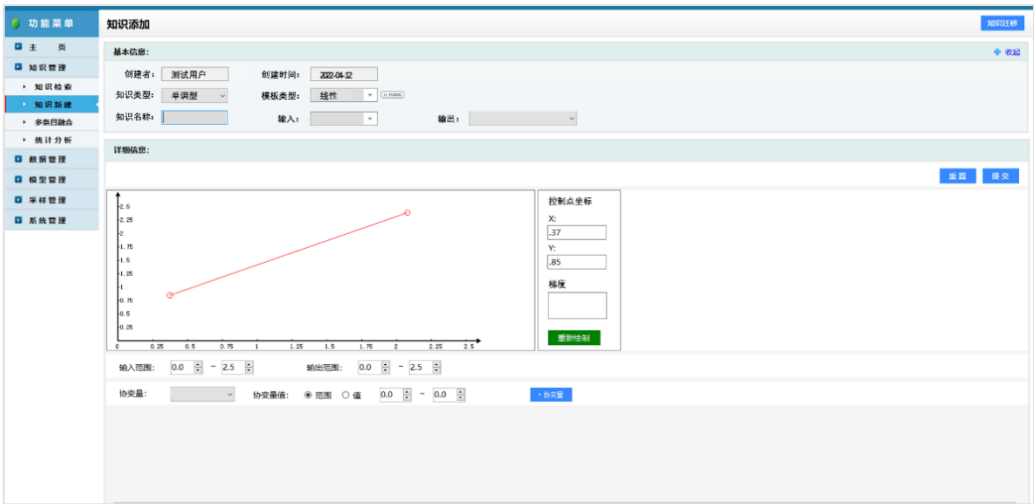


图 3.12 单调型知识新建系统页面

图 3.12 所示的为系统单调型知识的获取页面，这里用可拖动的直线来反映梯度的信息。

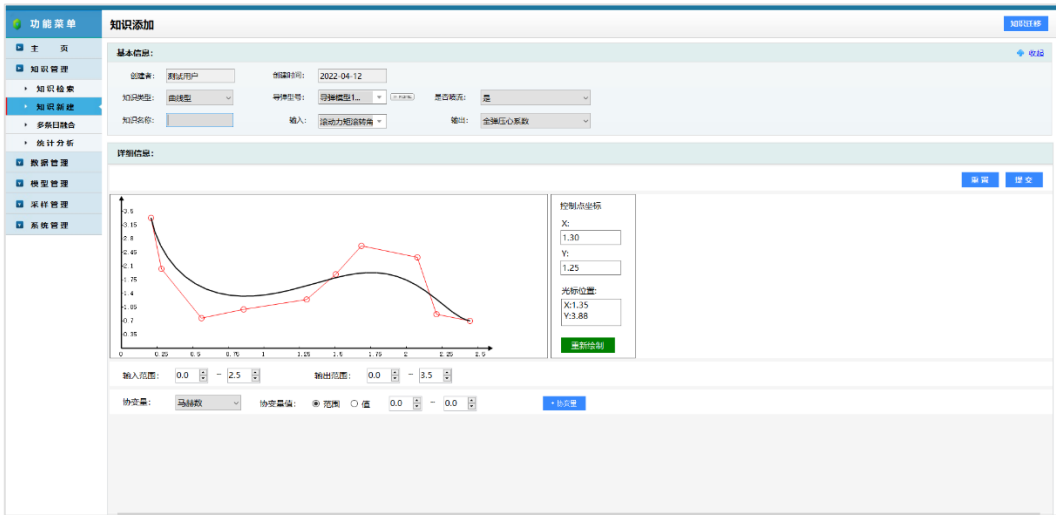


图 3.13 形状型知识新建系统页面

图 3.13 所示的为系统形状型知识的获取页面，这里在单调的直线基础上进一步实现控制点数量的自定义增减，可实现更复杂曲线关系的表征。

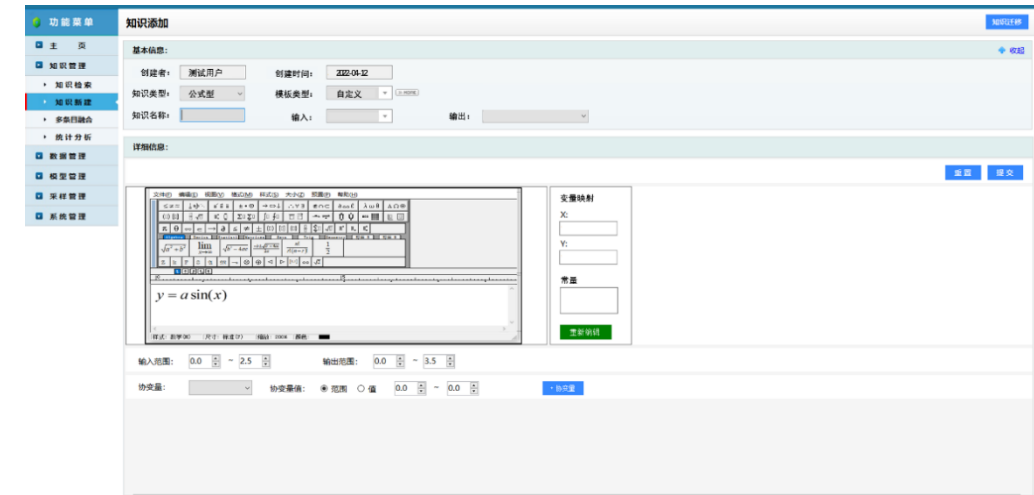


图 3.14 公式型知识新建系统页面

图 3.14 所示的为系统公式型知识的获取页面，通过调用 Mathtype 实现公式关系的编辑与修改。

3.4.2 工程知识的获取案例

为了进一步明确工程知识的获取方法，本节以某无人车桁架车身设计问题为例结合系统获取工具进行知识的获取。由于属性类型的知识主要是表征变量的取值范围，各类型的知识在系统工具中均包括该部分，不单独说明。本案例主要将存在的形状类型的工程知识获取过程进行详细说明。

首先某无人车桁架车身设计问题设计知识文本描述为：桁架车身结构主要由横向纵向肋梁组成，因此采用车身总长、总宽、总高、壳厚四个参数来表征车身整体参数。桁架体结构主要由三种不同边长和厚度的型材焊接而成，因此采用肋梁横截面的边长和厚度来表征局部体参数。其中性能目标包括：目标质量、最大位移和最大应力。知识涉及到厚度与最大应力之间的 s 型趋势关系，且最大应力随着厚度增加而降低。其中厚度取值范围为：2-6，最大应力取值范围：250-370，由于建模仿真需要一般采取归一化处理数据。

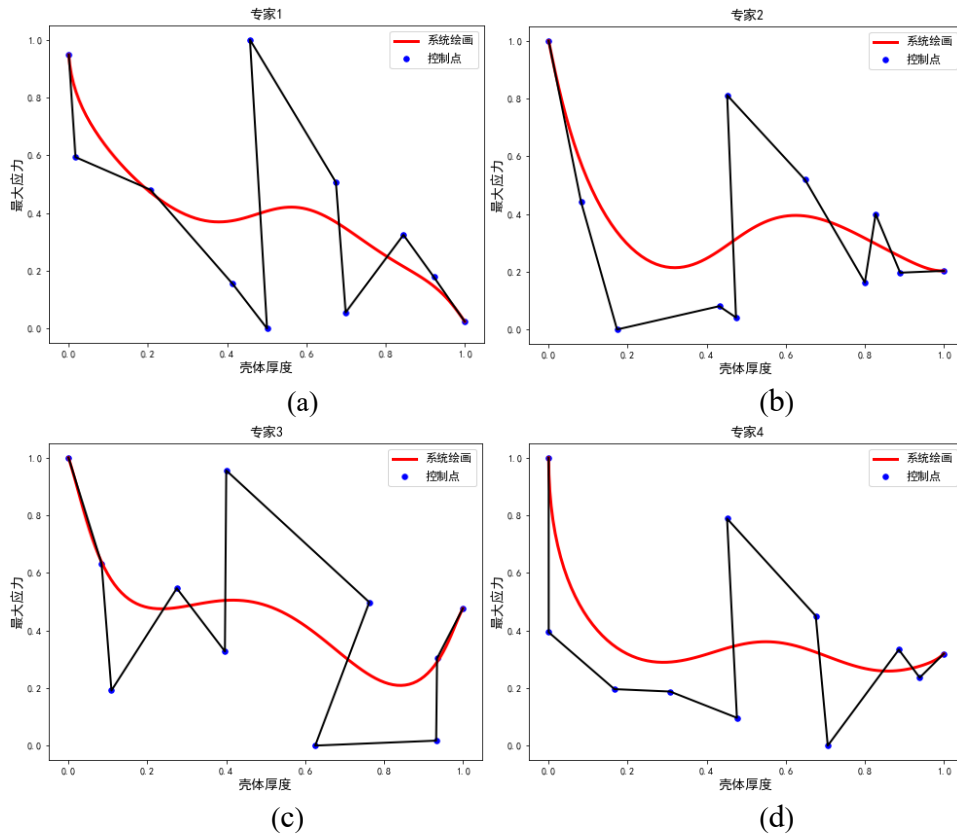
接下来针对该案例中涉及到工程知识获取步骤进行说明：

1. 模板匹配

通过知识文本进行模板匹配，分类的知识预设约束模板为形状-S 型（图 3.2 所示）、形状-直线型约束模板（图 3.5 所示）供专家进行选择。提取设计变量以及取值范围为：厚度[0,1]。性能目标：最大应力[0,1]。

2. 基于知识模板专家完成知识协同获取

基于形状-S 型约束模板，5 名设计专家参与知识修改。专家根据自身需求可获取多次，选择最符合自身认知的知识作为最终知识。图 3.15-(a)到图 3.15-(e)为各专家获取知识（仅展示知识曲线部分）：



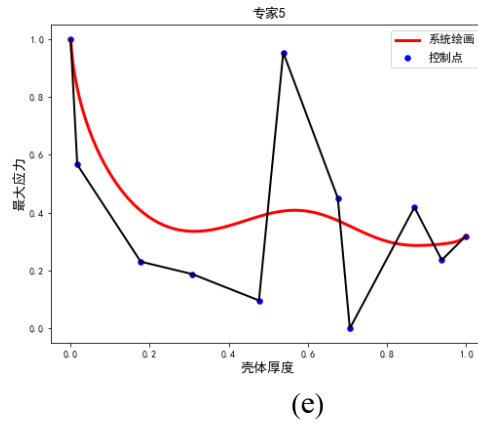


图 3.15 五名专家获取的工程知识

3.5 本章小结

本章提出了基于贝塞尔曲线的交互式工程知识获取技术,从而实现工程知识的快速获取。首先通过分析不同种类的工程知识的特点,完成了工程知识的定义与分类。再提出了基于文本匹配的方法实现知识预设模板的快速获取,最后针对不同类型的知识特点提出不同的交互式知识获取方法对知识模板进行协同修改以实现知识的获取。

第 4 章 基于多重一致性计算的工程知识检验技术

4.1 引言

工程知识的检验是为了将获取的知识中较其他有明显差异的知识筛选出来剔除掉，避免由于知识不一致而导致代理模型构建精度不升反降的问题。不同于公式、属性两种类型的工程知识具有准确的约束表达关系，形状类型工程知识的约束关系为抽象的曲线，由于受专家主观认知的影响，知识具有显著的个体差异。因此需要对形状类型的知识进行一致性检验，将差异较大的知识予以剔除以保证知识的信息趋于一致。

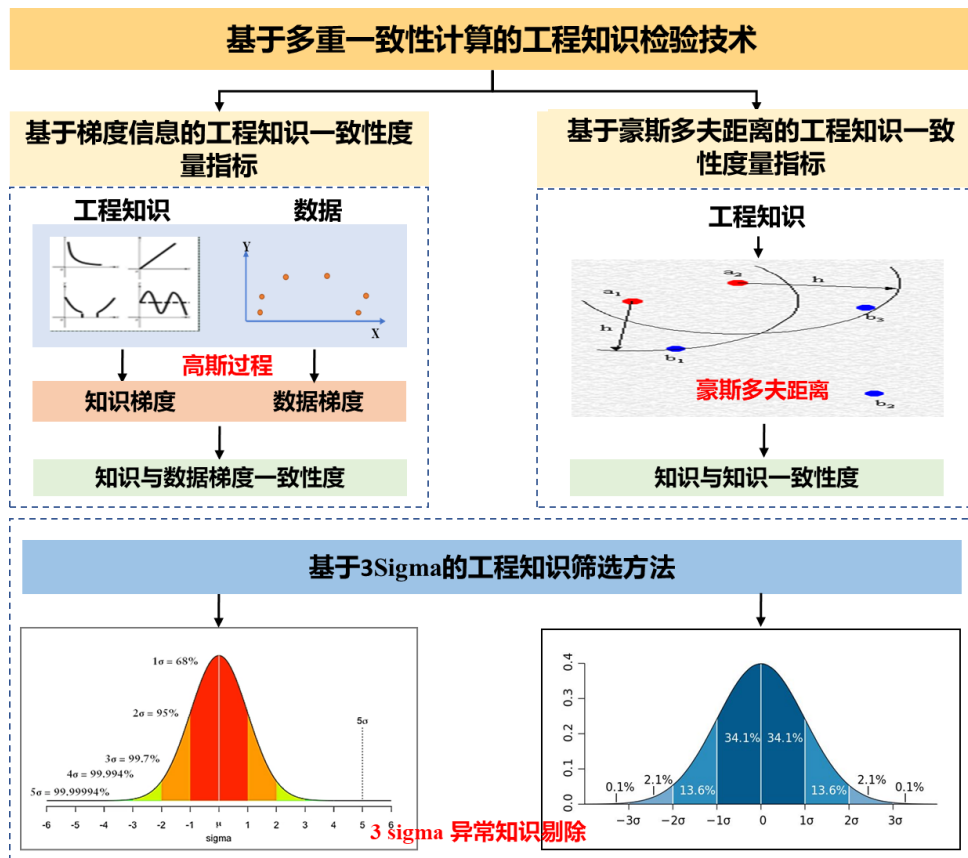


图 4.1 工程知识的一致性检验技术

融合知识构建高精度代理模型的核心是利用工程知识所表征的变量之间的映射信息去弥补数据样本覆盖设计空间不全面的缺陷。因此本章主要围绕带有专家主观认知的形状与单调型的知识进行检验与筛选。结合数据的准确性分别提出基于梯度信息的工程一致性度量指标和基于豪斯多夫距离的工程知识一致性度量指标。前者是构建代理模型的数据梯度信息作为参考对知识进行一致性度量，

后者则是针对知识之间整体相似水平作为评估知识是否一致的指标。最后利用3Sigma原则分析两种度量指标，完成知识的筛选。

4.2 知识一致性度量指标

知识与数据在融合知识构建代理模型中，数据与工程知识的目的均是让模型快速学习到变量之间的约束关系。这使得知识的一致性度量既可以从数据层面展开，又可以从自身展开。本文结合数据的精确程度提出两种检验思路：当可获取的数据精度高时，数据所体现的变量之间的约束关系准确，因此将其作为评估依据，提出基于梯度信息的工程知识一致性度量指标。当可用数据较少或数据精度不高时，则仅将知识之间的相似程度作为评估指标，提出基于豪斯多夫距离的工程知识一致性度量指标。

4.2.1 基于梯度信息的工程知识一致性度量指标

数据和知识在代理模型构建中都将融入模型中，这使得二者的一致性十分重要。因此本文利用梯度实现数据与知识的一致性度量。为了度量梯度一致性，拟将数据转换为高斯过程模型，进一步通过高斯过程模型求导获取梯度并与知识的梯度进行比较，分析二者的梯度一致性度。如图4.2所示，这部分的研究包括（1）基于数据信息的知识梯度提取；（2）知识梯度的获取；（3）计算梯度一致性度。

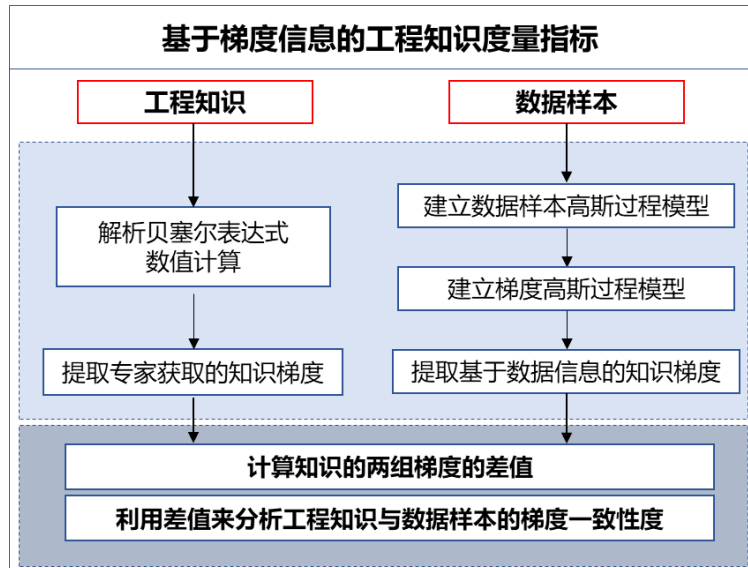


图 4.2 基于梯度的工程知识度量指标

4.2.1.1 基于高斯过程的知识梯度提取

复杂装备的设计问题往往具有维度高的特点,仅通过有限的数来构建代理模型难以实现满足要求的精度,具有不确定性。考虑到高斯过程本身具有天然的不确定性表征能力,可利用高斯过程模型对于有限的数进行建模,来提取潜在变量之间的梯度信息。首先论述高斯过程的基本原理。

➤ 高斯过程

首先,将 \mathbf{x} 定义为一个 D 维的设计变量, \mathbf{X} 为一个 $N \times D$ 的矩阵,其中包含了 N 组样本。

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^{(N-1)} \\ \mathbf{x}^N \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_d^1, \dots, \mathbf{x}_{(D-1)}^1, \mathbf{x}_D^1 \\ \mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_d^2, \dots, \mathbf{x}_{(D-1)}^2, \mathbf{x}_D^2 \\ \vdots \\ \mathbf{x}_1^{(N-1)}, \mathbf{x}_2^{(N-1)}, \dots, \mathbf{x}_d^{(N-1)}, \dots, \mathbf{x}_{(D-1)}^{(N-1)}, \mathbf{x}_D^{(N-1)} \\ \mathbf{x}_1^N, \mathbf{x}_2^N, \dots, \mathbf{x}_d^N, \dots, \mathbf{x}_{(D-1)}^N, \mathbf{x}_D^N \end{bmatrix} \quad (4.1)$$

高斯过程作为一种非参数化的贝叶斯预测模型,通常会假定满足零均值的先验分布:

$$p(\mathbf{f}|\mathbf{X}) = N(\mathbf{f}|\mathbf{0}, K(\mathbf{X}, \mathbf{X})) \quad (4.2)$$

式中, \mathbf{f} 为对应 \mathbf{X} 的预测值,其核心关键 $K(\mathbf{X}, \mathbf{X})$ 为协方差矩阵,取决于设计变量,通常由如公式 4.3 所示的平方指数协方差函数决定:

$$k(x_*, x_i) = \sigma_f^2 \exp\left(-\frac{1}{2}(x_* - x_i)^T \Lambda^{-1}(x_* - x_i)\right) \quad (4.3)$$

$$\Lambda = \begin{bmatrix} l_1^2 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & l_D^2 \end{bmatrix} \quad (4.4)$$

其中 σ_f 和 Λ 是高斯核函数超参, Λ 是对于 D 维的变量。高斯过程模型的预测 \mathbf{y} 可以视为基于实际性能 \mathbf{f} 的噪声观测值。由于计算机仿真模型并不能对于所有的细节进行准确地建模,一般预测值与实际的性能之间遵循高斯分布,如公式 4.5 所示:

$$p(\mathbf{y}|\mathbf{f}) = N(\mathbf{y}|\mathbf{f}, \sigma^2 \mathbf{I}) \quad (4.5)$$

式中, σ 是噪声的方差, \mathbf{I} 是单位矩阵,且 $\mathbf{y} = \mathbf{f} + \vartheta$, $\vartheta \sim N(0, \sigma^2)$,利用给定训练数据样本 \mathbf{X} 和 \mathbf{y} 构建好高斯过程后,对于测试点 \mathbf{x}_* 的条件预测分布为具有均

值和方差的高斯分布 $Y(\mathbf{x}_*) = N(\mu, \Sigma)$ ，相应函数表达式如下公式 4.6 和公式 4.7 所示：

$$\mu = K(\mathbf{x}_*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \quad (4.6)$$

$$\Sigma = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X}) \times (K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1} K(\mathbf{X}, \mathbf{x}_*) \quad (4.7)$$

► 高斯过程求导

结合公式 4.5 中噪声项 $\sigma^2 \mathbf{I}$ 的结构与协方差项 $K(\mathbf{X}, \mathbf{X})$ 相同，且其均不与 \mathbf{x}_* 相关，为方便公式撰写，现将 $(K(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})$ 定义为 $K(\mathbf{X}, \mathbf{X})$ ，将公式后部分 $K(\mathbf{X}, \mathbf{X})^{-1} \mathbf{y}$ 定义为 α 。则此时预测均值 μ 满足：

$$\mu = k(\mathbf{x}_*, \mathbf{X}) K(\mathbf{X}, \mathbf{X})^{-1} \mathbf{y} \triangleq k(\mathbf{x}_*, \mathbf{X}) \alpha \quad (4.8)$$

(1) 均值函数求导

首先分析协方差函数的导数：

$$\begin{aligned} \frac{\partial k(\mathbf{x}_*, \mathbf{x}_i)}{\partial \mathbf{x}_*} &= \frac{\partial}{\partial \mathbf{x}_*} \left\{ \sigma_f^2 \exp \left(-\frac{1}{2} (\mathbf{x}_* - \mathbf{x}_i)^T \Lambda^{-1} (\mathbf{x}_* - \mathbf{x}_i) \right) \right\} \\ &= \frac{\partial}{\partial \mathbf{x}_*} \left\{ -\frac{1}{2} (\mathbf{x}_* - \mathbf{x}_i)^T \Lambda^{-1} (\mathbf{x}_* - \mathbf{x}_i) \right\} k(\mathbf{x}_*, \mathbf{x}_i) \\ &= -\Lambda^{-1} (\mathbf{x}_* - \mathbf{x}_i) k(\mathbf{x}_*, \mathbf{x}_i) \end{aligned} \quad (4.9)$$

进一步将训练点扩展为整个训练集合，定义 $\tilde{\mathbf{X}}_* = [\mathbf{x}_* - \mathbf{x}_1, \dots, \mathbf{x}_* - \mathbf{x}_N]^T$ ，此时针对测试点 \mathbf{x}^* ，其预测均值一阶导数满足：

$$\bar{\mu} = \frac{\partial k(\mathbf{x}_*, \mathbf{X})}{\partial \mathbf{x}_*} = -\Lambda^{-1} \tilde{\mathbf{X}}_*^T (k(\mathbf{x}_*, \mathbf{X})^T \odot \alpha) \quad (4.10)$$

(2) 方差函数求导

根据上部分可计算得到后验均值的导数，这里从两个测试点分析其方差导数关系，假设：

$$Y(\mathbf{x}_*) = \mu(\mathbf{x}_*) + z_* \quad (4.11)$$

$$Y(\mathbf{x}_* + \delta) = \mu(\mathbf{x}_* + \delta) + z_\delta \quad (4.12)$$

同时满足：

$$P(z_*, z_\delta) = N \left(0, \begin{bmatrix} k_{**} - k_*^T K^{-1} k_* & k_{*\delta} - k_*^T K^{-1} k_\delta \\ k_{\delta*} - k_\delta^T K^{-1} k_* & k_{\delta\delta} - k_\delta^T K^{-1} k_\delta \end{bmatrix} \right) \quad (4.13)$$

此时，导数满足：

$$\begin{aligned}
\frac{\partial Y(x_*)}{\partial x_*} &= \lim_{\delta \rightarrow 0} \frac{\mu(x_* + \delta) - \mu(x_*)}{x_* + \delta - x_*} \\
&= \lim_{\delta \rightarrow 0} \frac{\mu(x_* + \delta) + z_\delta - \mu(x_*) - z_*}{\delta} \\
&= \lim_{\delta \rightarrow 0} \frac{\mu(x_* + \delta) - \mu(x_*)}{\delta} + \lim_{\delta \rightarrow 0} \frac{z_\delta - z_*}{\delta} \\
&= \bar{\mu}(x_*) + \lim_{\delta \rightarrow 0} \frac{z_\delta - z_*}{\delta}
\end{aligned} \tag{4.14}$$

式子前半部分为均值函数的导数，后半部分则为方差的导数，进一步分析：

$$\begin{aligned}
\mathbf{V} \left[\lim_{\delta \rightarrow 0} \frac{z_\delta - z_*}{\delta} \right] &= \lim_{\delta \rightarrow 0} \frac{1}{\delta^2} (\mathbf{V}[z_\delta] + \mathbf{V}[z_*] - C[z_\delta, z_*] - C[z_*, z_\delta]) \\
&= \lim_{\delta \rightarrow 0} \frac{1}{\delta^2} (k_{\delta\delta} - k_\delta^T K^{-1} k_\delta + k_{**} - k_*^T K^{-1} k_* \\
&\quad - (k_{*\delta} - k_*^T K^{-1} k_\delta) - (k_{\delta*} - k_\delta^T K^{-1} k_*)) \\
&= \lim_{\delta \rightarrow 0} \frac{1}{\delta^2} (k_{\delta\delta} - k_{*\delta} - k_{\delta*} + k_{**} \\
&\quad - (k_\delta - k_*)^T K^{-1} (k_\delta - k_*)) \\
&= \frac{\partial^2 k(x_*^i, x_*^j)}{\partial x_*^i \partial x_*^j} - \frac{\partial k(x_*^i, X)}{\partial x_*^i} K^{-1} \frac{\partial k(X, x_*^j)}{\partial x_*^j}
\end{aligned} \tag{4.15}$$

此时， x_*^i, x_*^j 代表 x_* 中的 i, j 维度，整理可以得到关于测试点 x_* ，其一阶梯度预测值满足：

$$\mathbf{P} \left(\frac{\partial Y(x_*)}{\partial x_*} \right) = N \left(\frac{\partial k(x_*, X)}{\partial x_*}, \frac{\partial^2 k(x_*^i, x_*^j)}{\partial x_*^i \partial x_*^j} - \frac{\partial k(x_*, X)}{\partial x_*} K^{-1} \frac{\partial k(X, x_*)}{\partial x_*} \right) \tag{4.16}$$

➤ 梯度信息提取

针对有限数据进行高斯过程建模，将所获取知识的设计变量作为特征维度输入，其他维度保持不变。例如工程知识涉及的是第 j 维的设计变量与性能目标之间的关系，则将第 j 维的数据更换为工程知识的输入区间，其他维度保持在该条知识下原有的取值。结合高斯过程求导原理实现基于数据信息的工程知识的梯度的预测，预测结果包括均值与方差：

$$g_{GP} = \mathbf{P} \left(\frac{\partial Y(B_{n(t)_x})}{\partial B_{n(t)_x}} \right) \tag{4.17}$$

其中， $B_{(t)}^x, B_{(t)}^y$ 分别是拟合曲线上点 $B_{(t)}$ 的横、纵坐标

4.2.1.2 工程知识梯度的获取

根据贝塞尔曲线解析式，拟合曲线上的观测点本身由控制点的坐标跟 t 的采样密度决定，因此拟合曲线上的观测点的横纵坐标均由控制点的横纵坐标定义。这里根据链式求导法则，利用 t 作为中间变量，计算拟合曲线上观测点的导数值，如下公式 4.18 所示：

$$g_{know} = \frac{\partial B_{(t)}^y}{\partial B_{(t)}^x} = \frac{\frac{\partial B_{(t)}^y}{\partial t}}{\frac{\partial B_{(t)}^x}{\partial t}} \quad (4.18)$$

$$= \frac{\sum_{i=0}^n C_n^i P_i^y (1-t)^{n-i-1} t^{i-1} (i-nt), t \in [0,1]}{\sum_{i=0}^n C_n^i P_i^x (1-t)^{n-i-1} t^{i-1} (i-nt), t \in [0,1]}$$

其中 P_i^x 、 P_i^y 指的是第 i 控制点的横、纵坐标，最终可以得到 t 组梯度。

4.2.1.3 度量指标定义

高斯过程模型针对拟合曲线上某点的预测结果包括均值(μ)和标准差(σ)， $1.96 * \sigma$ 表示在该均值的基础上下浮动的 95% 的置信度水平， σ 值越小表明预测的不确定性范围越小，预测值越准确。本文为充分考虑不确定性范围即 σ 值的影响，引入一个误差的惩罚项 w_i 如公式 4.19 所示。当高斯过程求导预测的均值与各知识获取的梯度值相差大于 $1.96 * \sigma$ 时，需要额外乘上一个误差系数，反之则不用。目的是当预测值在 95% 置信水平内时，不额外考虑方差带来的影响。

$$w_i = \begin{cases} 1 & 1.96 * \sigma_i \geq |\mu_i - g_{know}^i| \\ 1 + \frac{1}{1 + e^{-\frac{1.96 * \sigma_i}{|\mu_i - g_{know}^i|}}} & 1.96 * \sigma_i < |\mu_i - g_{know}^i| \end{cases} \quad (4.19)$$

式中， μ_i 、 σ_i 分别为高斯过程模型针对工程知识某输入的预测均值与方差， g_{know}^i 是基于各知识对应某输入得到的梯度值。紧接着将获取的两组梯度值进行一致性度量，这里不仅仅考虑梯度数值的大小差异，就其梯度正负性也做了度量。

误差一致性：基于均方根误差（Root Mean Square Error, RMSE），结合惩罚项 w_i 来定义误差一致性：

$$RMSE_p = \sqrt{\frac{\sum_{i=1}^N (\mu_i - g_{know}^i)^2 * w_i}{N}} \quad (4.20)$$

符号一致性：如公式 4.21 所示计算二者梯度符号一致的数量占比， N 为对工程知识进行梯度预测的输入数量， n_c 为两者梯度符号一致的数量。这里存在特殊情况需要说明：即高斯过程的预测均值与知识梯度值不一致时，若二者的差值小于 $1.96 * \sigma$ 时，仍认定为二者符号一致，反之则认定不一致：

$$symbolic\ consistency = \frac{n_c}{N} \quad (4.21)$$

最终指标：为最大化差异一致性指标之间的误差，这里将误差一致性值与符号一致性值的商作为最终的一致性度量指标。考虑到符号一致性值是一个小于 1 的值，这里是对误差的一致性值做进一步的放大，增大知识之间的指标差异，最终得到知识与样本的一致性度量指标 g_{final} ：

$$g_{final} = \sqrt{\frac{N * \sum_{i=1}^N [(\mu_i - g_{know}^i)^2 * w_i]}{n_c^2}} \quad (4.22)$$

4.2.2 基于豪斯多夫距离的工程知识一致性度量指标

单调型的知识作为一种特殊的形状曲线，通过贝塞尔曲线的获取方法能够基于直线进行表征。这使得知识之间的一致性可统一由知识对应曲线相似度来分析。结合豪斯多夫距离具有度量曲线形状近似程度的能力，本文选择豪斯多夫距离作为知识之间一致性度量的依据，计算各知识与其他所有知识的豪斯多夫距离的均值来定义知识的整体性相似水平。如图 4.3 所示，这部分的研究包括（1）豪斯多夫距离；（2）度量指标。

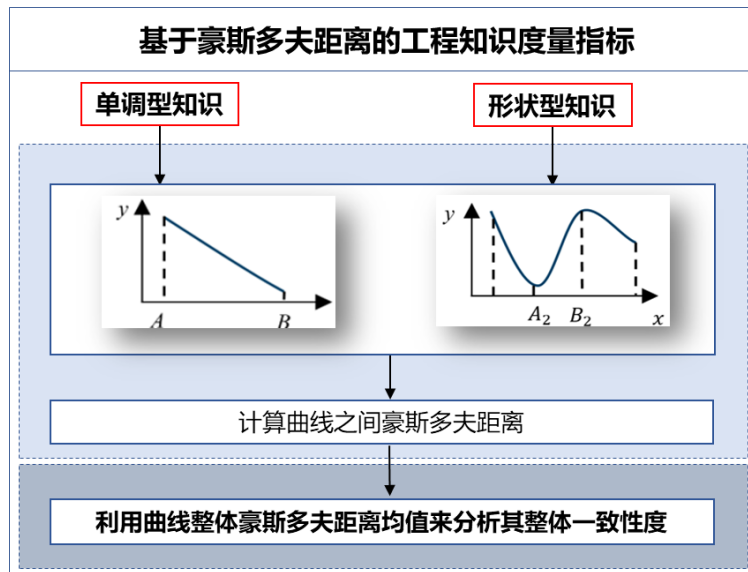


图 4.3 基于豪斯多夫距离的度量指标

4.2.2.1 豪斯多夫距离

豪斯多夫距离（Hausdorff Distance）是匹配点特征的一种方法，它不需要建立点与点之间的一一对应关系，只需要计算两个真子集之间的相似程度最大距离，能有效地处理多特征点的情况。真子集可以是一组有限的点集，因此豪斯多夫距离可以描述成一个点集中的点到另一个点集中所有点的最短距离的最大值^[75]。

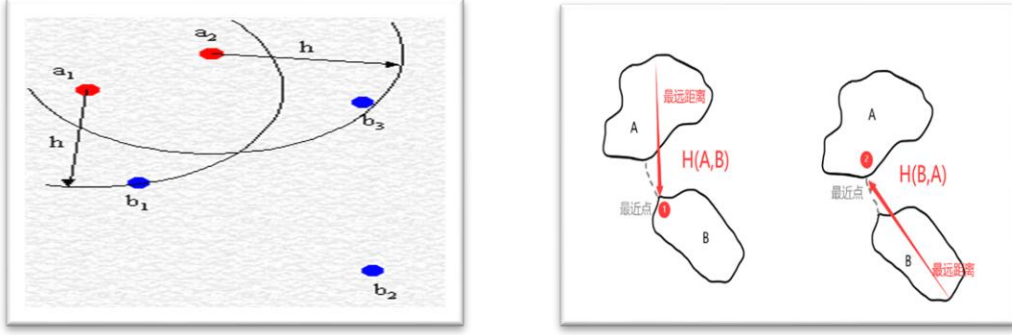


图 4.4 豪斯多夫距离示意图

豪斯多夫距离的数学描述如下：

如图 4.4 所示，假设空间中存在两个点集 $A = \{a_1, a_2, \dots, a_p\}$ ， $B = \{b_1, b_2, \dots, b_q\}$ ，则集合 A ， B 间的豪斯多夫距离定义为：

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (4.23)$$

$H(A, B)$ 为 A 集合与 B 集合间的双向豪斯多夫距离，其中 $h(A, B)$ ， $h(B, A)$ 分别为 A 集合到 B 集合和 B 集合到 A 集合的单向豪斯多夫距离，两者数学定义式如下：

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (4.24)$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \|a - b\| \quad (4.25)$$

其中 $\|\cdot\|$ 是 A 集合到 B 集合间的距离范式。 A 集合与 B 集合之间的豪斯多夫距离可以理解为一个点集中的点到另一个点集的最短距离的最大值，它度量了 A 和 B 两个点集间的最大不匹配度^[76]。两个点集之间的豪斯多夫距离越小，表明二者最大不匹配程度越小，相似程度越高。本文正是利用这一点，将工程知识统一转换为曲线形式，利用曲线上点的集合进行豪斯多夫距离度量。最终通过该指标来度量知识之间的相似性程度来分析知识之间的一致性度。

4.2.2.2 度量指标定义

两条知识之间的一致性度量可通过两者之间的豪斯多夫距离表征，表征知识的整体相似度水平则通过公式 4.26 所示：

$$g_{final} = \frac{1}{n-1} \sum_{j=1}^{n-1} H(Exp_i, Exp_j) \quad (4.26)$$

式中， n 是知识的数量， Exp_i 为工程知识转换成的曲线（点集），单调型与形状型知识的点集直接选择贝塞尔曲线拟合出的系列点 $B_{(t)}$ 。 $H(Exp_i, Exp_j)$ 指的是计算二者之间的豪斯多夫距离。公式的核心在于计算某条知识较其他剩余的所有知识点集的豪斯多夫距离均值，通过较其他知识的平均相似度来定义其整体一致性度。

4.3 基于 3Sigma 的工程知识筛选方法

结合 4.2 节以及 4.3 节提出的两种知识度量指标，知识的一致性度可以通过两种指标进行定义计算。因此，对于知识异常检验可以转换成对于知识一致性度的异常值检验。3Sigma 原则常作为异常检验中的重要原则，本文在此基础上提出基于 3Sigma 的工程知识筛选方法。首先论述 3Sigma 原则的基本原理。

4.3.1 3Sigma 原则

3Sigma 原则（3Sigma 原则）是常用的误差判别原则， σ 代表标准差， μ 代表均值，3Sigma 原则为：

数值分布在 $(\mu - 1\sigma, \mu + 1\sigma)$ 中的概率为 0.6826；

数值分布在 $(\mu - 2\sigma, \mu + 2\sigma)$ 中的概率为 0.9544；

数值分布在 $(\mu - 3\sigma, \mu + 3\sigma)$ 中的概率为 0.9974；

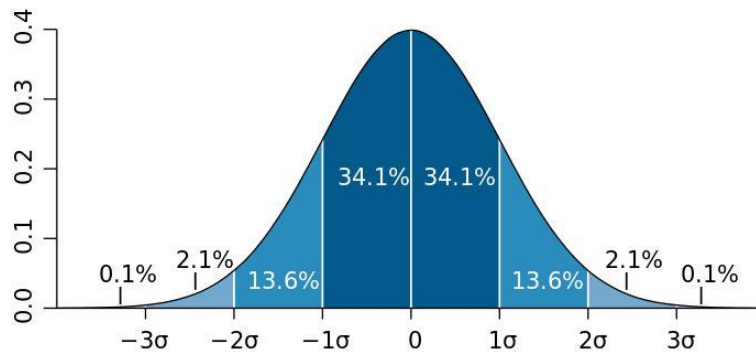


图 4.5 3Sigma 原则

可以认为,取值几乎全部集中在 $[\mu - 3\sigma, \mu + 3\sigma]$ 区间内,超出这个范围的可能性仅占不到 0.3%,因此将超出的取值作为异常取值予以剔除。

4.3.2 知识筛选步骤

本文在结合两种知识检验指标 g_{final} (公式 4.22 和公式 4.26),提出了基于 3Sigma 原则的异常知识筛选方法,主要步骤如下:

步骤 1: 针对所有知识的 g_{final} 指标,计算均值 g_{μ} 与标准差 g_{σ} ,形成一个概率分布 $N(g_{\mu}, g_{\sigma}^2)$ 。

步骤 2: 按照 3Sigma 原则实现对异常知识的剔除。这里选择大于 3Sigma 的知识予以剔除,即当某知识指标满足 $|g_{final} - g_{\mu}| \geq 3g_{\sigma}$,认定该知识为异常知识予以剔除。

基于两种不同场景下的知识一致性度量指标,结合 3Sigma 原则,本文实现了针对形状、单调类型这两类知识的筛选过程。

4.4 实验设计

4.4.1 实验目的

本章需要探究的实验目的包括:(1) 基于两种知识检验方法在处理多知识时,能否有效实现异常知识的剔除。(2) 结合两种方法的应用场景,分析在不同的数据精度背景下,两种知识检验方法的适用性。

4.4.2 实验内容

通过 Matyas、Branin 测试函数以及无人车桁架车身设计案例对实验目的展开验证过程。两个测试函数为常用的基准测试函数(Benchmark Function)之一,可在维基百科上找到,实验涉及的均是形状型知识。

4.4.2.1 测试函数

➤ Matyas 函数

Matyas 测试函数常用于测试的二维函数,表达式为:

$$f(x) = 0.26(x_1^2 + x_2^2) + 0.48x_1x_2 \quad (4.27)$$

其中 $x_1 \in [-10,10], x_2 \in [-10,10]$ 。实验获取的知识为: $x_2 = 0.5$ 时, $f(x)$ 关

于 x_1 的曲线关系，其中 $x_1 \in [-10, 10]$, $f(x) \in [0, 30]$ 。如图 4.6 所示为基于贝塞尔方法绘制得到的基准曲线，曲线形状简单但是曲线坐标覆盖的范围广。实验中用以构建高斯过程的数据为区间内均匀采样 400 组数据。

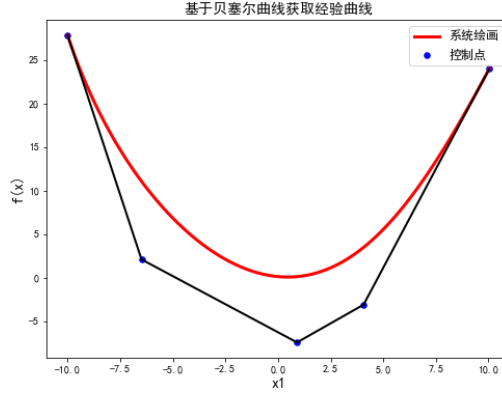


图 4.6 Matyas 函数基准曲线

➤ Branin 函数

Branin 作为一个常用的测试函数，表达式为：

$$f(x) = a(x_2 - bx_1^2 + cx_1 - r)^2 + s(1 - t) \cos(x_1) + s \quad (4.28)$$

其中 $a = 1, b = \frac{5.1}{4\pi^2}, c = \frac{5}{\pi}, r = 6, s = 10, t = \frac{1}{8\pi}, x_1 \in [-5, 10], x_2 \in [0, 15]$

实验获取的知识为： $x_2 = 5$ 时， $f(x)$ 关于 x_1 的曲线关系，其中 $x_1 \in [-5, 10], f(x) \in [5, 160]$ 。如图 4.7 所示为基于第二节知识获取方法得到的基准曲线，曲线形状复杂且曲线坐标覆盖的范围广。实验中用以构建高斯过程的数据为区间内均匀采样 400 组数据。

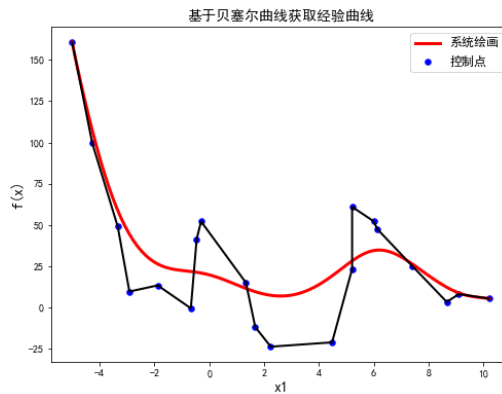


图 4.7 Branin 函数基准曲线

4.4.2.2 工程案例

无人车桁架车身设计由于任务要求高,面临多工况环境,如图 4.8 所示为无人车桁架车身结构模型,车身结构参数近百个(包括全局整体参数、局部加固及梁结构参数及定位参数)。为了简化设计初期的模型,本文只考虑关键设计参数。

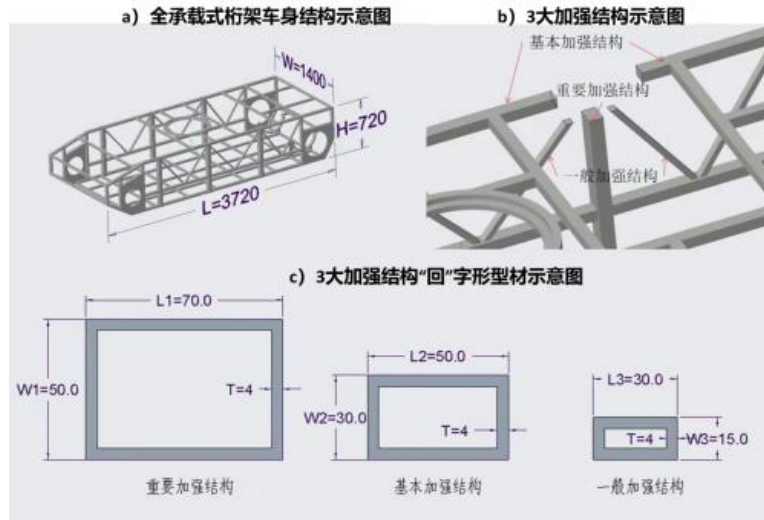


图 4.8 无人车桁架车身结构模型

由于桁架车身结构主要由横向纵向肋梁组成,因此采用车身总长、总宽、总高、壳厚四个参数来表征车身整体参数(图 4.8-a)。桁架体结构主要由三种不同边长和厚度的型材焊接而成,因此采用肋梁横截面的边长和厚度来表征局部体参数(图 4.8-b、4.8-c)。该案例获取的知识曲线围绕壳体厚度与最大应力性能之间展开,经归一化处理后壳体厚度与最大应力范围均在 $[0,1]$ 。知识曲线选择第三节案例中专家 5 获取的知识曲线为基准曲线(图 4.9 所示),曲线形状复杂但曲线坐标覆盖的范围小。实验中用以构建高斯过程的数据为实验仿真环境下获取的 400 组数据。

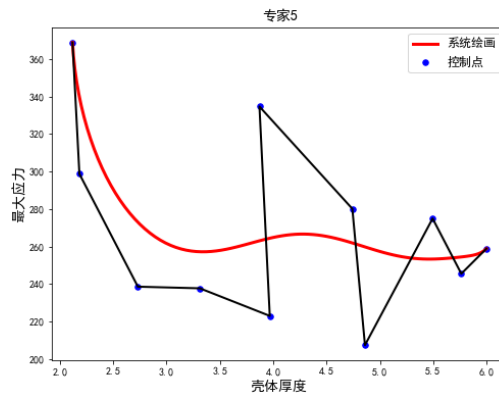


图 4.9 无人车桁架车身案例最大应力与壳体厚度知识基准曲线

4.4.2.1 实验步骤

步骤 1: 知识生成。在基于贝塞尔曲线的知识获取方法中，知识曲线是由一组控制点的坐标确定，不同控制点的坐标对应不同的形状。因此，实验基于上述案例基准知识曲线，通过在控制点上加噪声来形成不同的知识曲线以获取知识样本。添加的噪声为高斯噪声 $n * N(0,1) * base$ ， $base$ 指的是基准曲线控制点横、纵坐标的标准差以区分各实验案例的差异，其中 n 依次取值 0.5、1、2，生成知识曲线占比分别为 20%、30% 和 50%。样本共取值重复 20 轮，每轮生成 50 条知识曲线。

步骤 2: 知识筛选。针对每个案例分别按照基于梯度信息的工程知识一致性度量方法、基于豪斯多夫距离的工程知识一致性度量方法对每一组 50 条知识进行度量，利用 3Sigma 原则进行异常知识的筛选。实验对照组选择保留所有的知识曲线不进行过滤筛选。

步骤 3: 误差检验。误差检验指标结合均方根误差跟豪斯多夫距离进行定义，表达式如公式 4.29 所示：

$$RMSE_H = \sqrt{\frac{\sum_{i=1}^M (H(Exp_{base}, Exp_i))^2}{M}} \quad (4.28)$$

其中， M 为筛选剩余知识的数量， $H(Exp_{base}, Exp_i)$ 计算基准线与筛选后某知识曲线的豪斯多夫距离。

步骤 4: 重复步骤 2-3。完成 20 轮实验，计算 20 组预测误差的平均值作为案例的最终误差。（目的 1）

步骤 5: 重复步骤 2-4。其中针对基于梯度信息的知识度量方法，对用于构建高斯过程的数据增加高斯噪声以降低数据精度进行对照试验，同步骤 4 中两种知识筛选方法进行对比分析。（目的 2）

4.5 实验结果与分析

三个案例的最终预测结果误差如表 4.1 所示，表中的实际误差值是实验运行 20 轮后，每组实验中筛选剩余的知识曲线与基准曲线误差的平均值，目的是减少方法本身跟随机偶然性对实验结果带来的影响。在表中，“梯度”表示基于梯度信息的知识度量指标，“带噪声梯度”表示在基于梯度指标用于建模的数据

添加了噪声，“豪斯多夫距离”表示基于豪斯多夫距离的度量指标，“无过滤”表示未对知识进行筛选保留所有的知识曲线。

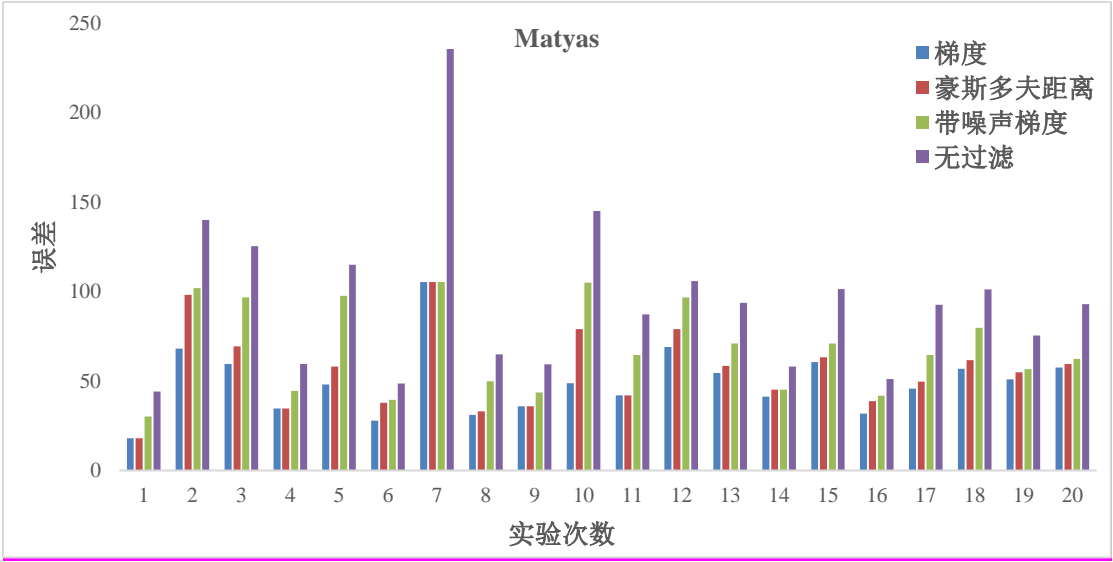
表 4.1 知识检验实验结果

实验结果				
检验指标	梯度	带噪声梯度	豪斯多夫距离	无过滤
Matyas	49.37	68.37	56.11	94.82
Branin	210.41	396.27	322.15	572.75
车架	0.86	1.12	0.91	1.47

由表中可以发现：

（1）三种知识检验指标的误差相较于保留所有的知识方法明显更低，表明均能够剔除掉差异较大的知识曲线。

（2）在数据未带噪声的情形下，基于梯度信息的度量指标误差优于基于豪斯多夫距离均值指标。然而在数据带有噪声情形下，后者的预测误差明显好于前者。这也表明，在数据精度高时，基于梯度信息的筛选效果更好；反之则是基于豪斯多夫距离的度量指标更优。由图 4.10 可以发现，各案例的 20 轮实验每组的误差对比，上述结论也很明显。



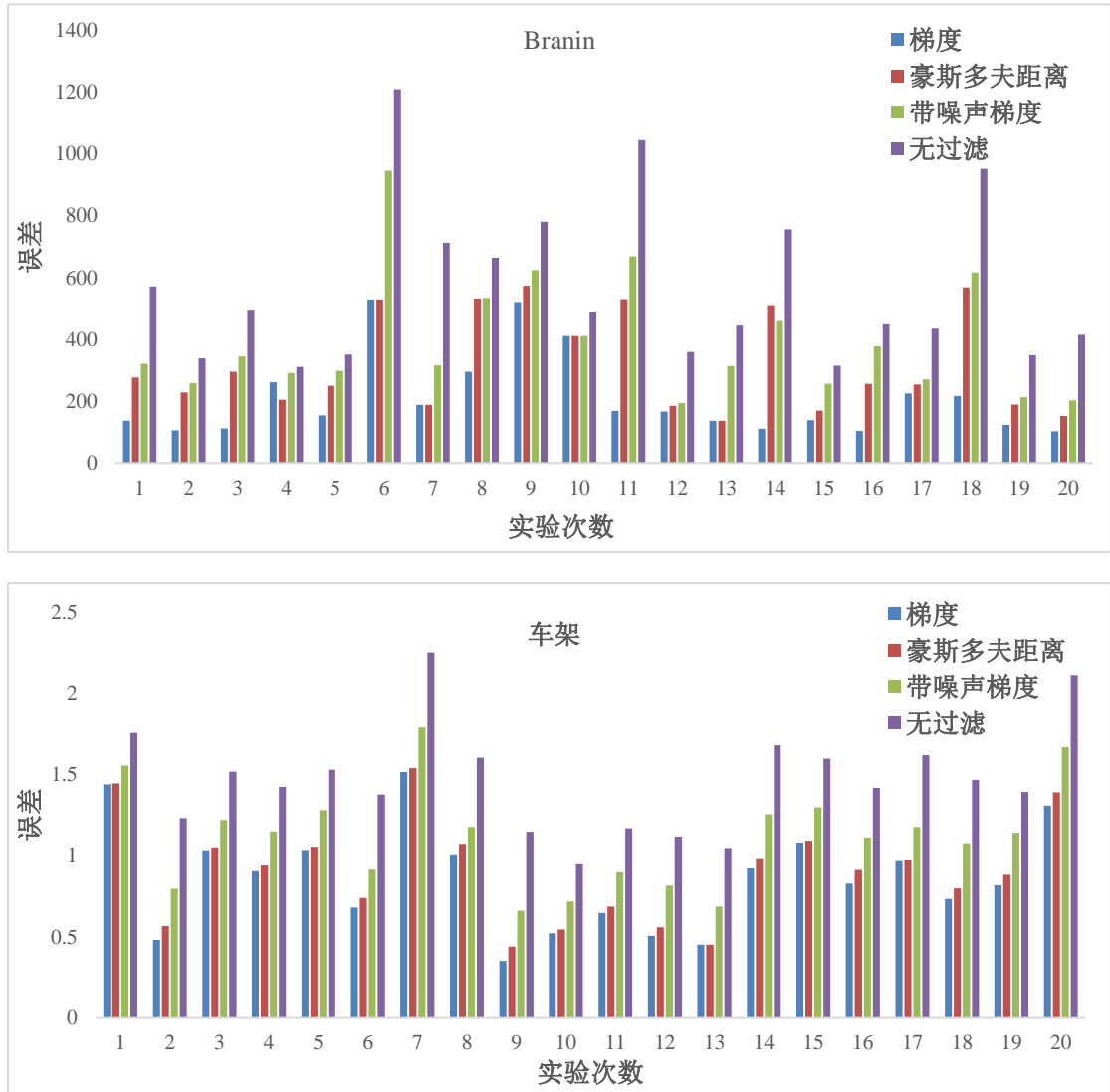


图 4.10 各案例 20 次运行的每次误差值

4.6 本章小结

本章节主要围绕工程知识检验与筛选展开研究,分别提出了基于梯度信息的工程知识一致性度量指标、基于豪斯多夫距离的工程知识一致性度量指标,通过两种指标定义知识的一致性度,结合 3Sigma 原则对知识的一致性度进行异常值检验来实现异常知识的筛选。最后通过实验验证两种工程知识检验技术均能有效剔除异常的知识,但是基于梯度信息的工程知识一致性检验方法较豪斯多夫距离的检验方法在数据精度高的情形下有更好的检验效果,这使得两种检验方法对应着数据精度的不同情形。

第 5 章 基于高斯过程的多条目工程知识融合技术

5.1 引言

经过两种指标对多专家协同获取的工程知识进行度量筛选后, 尽管异常的知识会被剔除, 保留的知识数量仍然众多。众多的工程知识直接融合进代理模型中会提升模型计算复杂度, 影响代理模型收敛效果。这使得在多知识曲线基础上找到的一条融合曲线, 既能显著降低知识数量的同时还涵盖了知识的有效信息。如图 5.1 所示, 本文从曲线整体、曲线上个体点两个层面提出不同的知识融合策略实现多知识的融合, 最后基于高斯过程模型对融合知识曲线进行建模, 实现多知识统一的概率模型表征。

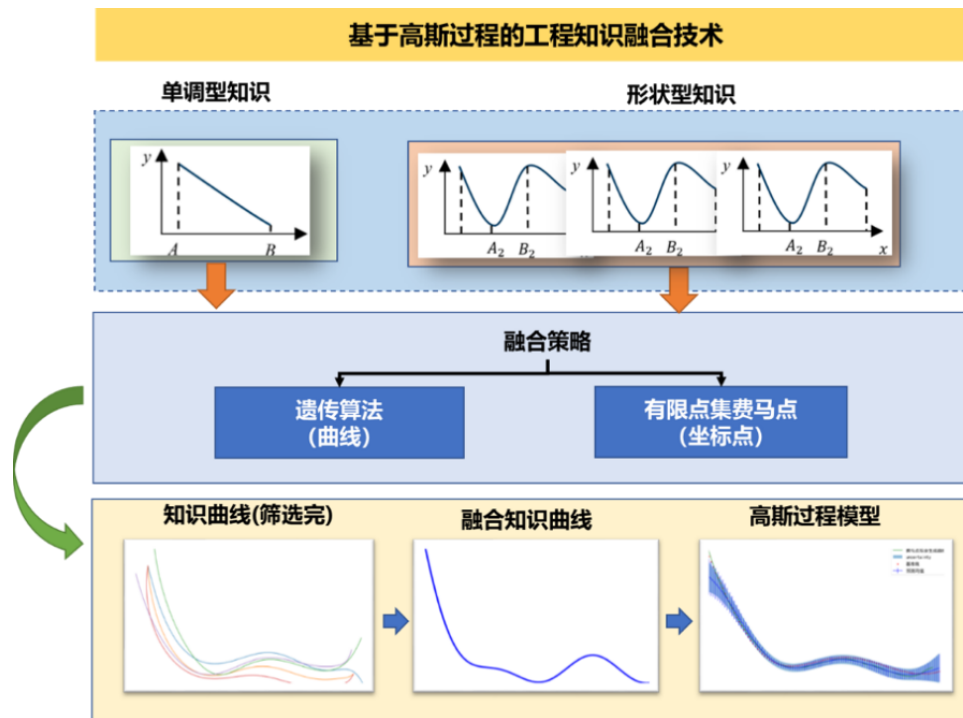


图 5.1 工程知识的融合

5.2 基于遗传算法的曲线信息融合方法

遗传算法 (Genetic Algorithm, GA) 起源于对生物系统所进行的计算机模拟研究。它是模仿自然界生物进化机制发展起来的随机全局搜索和优化方法, 可自适应地控制搜索过程以求得目标条件的最佳解。本方法就是利用遗传算法, 将融合知识曲线与各知识曲线的豪斯多夫距离均值作为寻优目标, 自适应求得与各知

识整体相似度最高的融合知识。首先论述遗传算法的基本流程。

5.2.1 遗传算法

遗传算法的主要特点为全局性、并行性、高效性。它通过编码技术来模拟自然生物遗传过程中群众的演变过程，利用新个体的复制、交叉、变异来实现种群进化，以此来实现搜索的全局性。遗传算法在搜索的时候是多个途径同时进行的，这使得算法的计算能力较高。此外，遗传算法研究起步早，拥有完整的研究体系，目前已广泛应用于故障诊断、数据挖掘、图像处理等领域。

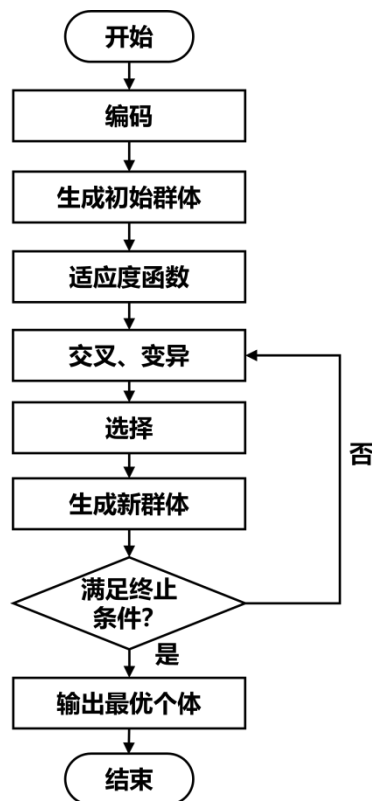


图 5.2 遗传算法基本流程

如图 5.2 所示为遗传算法运算流程，主要的步骤如下：

(1) 编码 (Coding & Decoding)

编码 (Coding) 的主要目的是将计算机无法理解的现实问题的表现类型转换为计算机既可以理解又可以实现遗传进化的基因型。不同的编码方式对于后续种群进化过程的交叉与变异的步骤有着不同的影响，因此选择合适的编码方式极大程度影响着算法效率。解码 (Decoding) 则是与编码的流程相反，它是将编码的基因形式转换成描述现实问题的表征类型。目前主流的编码方式有二进制编码、

格雷码编码、浮点数（实数）编码方式、排列编码和符号编码等^[77]。

（2） 适应度评估（Evaluation）

个体评估的核心就是适应度计算，它是利用适应度函数对个体的优劣进行评价，一般适应度函数可直接由目标函数转换得到。

（3） 遗传操作（Reproduction）

遗传操作的核心是模拟种群进化中不断产生新个体的过程，遗传基因 DNA 通过复制而转移到新个体中，新个体就继承了上一代的基因。遗传主要分为交叉（Crossover）、变异（Mutation）两种方法进行：

交叉：交叉指不同个体的 DNA 配对并通过某种方式对染色体中的部分基因段进行互相交换，生成新的个体 DNA。

变异：变异是指将个体中 DNA 的某一基因段用符合相应位置要求的其他的基因代替，从而构成新的基因编码来产生新的个体 DNA。

（4） 选择（Selection）

选择是用来确定如何从父代与子代中选择个体形成新的种群的方法，以便优秀的个体 DNA 在接下来的进化中得以保留。目前主流的选择方法包括轮盘赌、随机竞争选择、均匀排序、确定式选择、最佳保存策略、排挤选择等^[78]。

（5） 终止判断

经过以上步骤，种群不断的遗传产生新的个体，并通过优胜劣汰机制始终保留优质个体。进化过程中需要对新种群进行评估以判断是否终止，具体判断条件有两个：迭代代数和适应度值。终止的判断过程为：如果遗传代数达到迭代代数，则停止；如果个体适应度值达到预期目标，则停止。否则，进行新一轮的交叉变异操作，直到满足终止条件。

5.2.2 基于遗传算法的曲线信息融合方法

本方法将曲线由一组控制点经过贝塞尔公式拟合（公式 3.4）来定义，并通过个体 DNA 表征控制点坐标进行遗传操作以模拟不同的融合曲线。同时，将融合知识曲线与各知识曲线的豪斯多夫距离均值作为寻优目标，自适应求得与各知识的整体相似度最高的融合曲线。整体算法流程如图 5.3 所示：

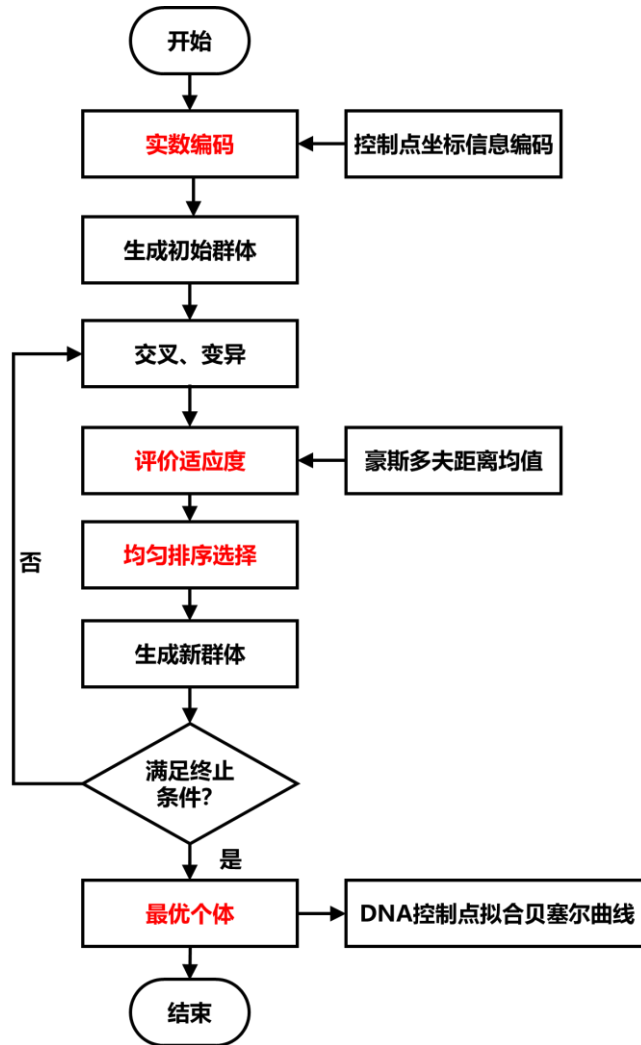


图 5.3 基于遗传算法的曲线信息融合方法流程

步骤 1：初始化

确定初始种群数目 $N=200$ ，随机产生初始种群 pop 。初始种群中个体是一组二维的坐标点，结合贝塞尔公式（公式 3.20）可将个体 DNA 拟合成一条知识曲线。

步骤 2：编码方式

为了避免二进制编码和译码繁琐、计算量大等特点，个体的编译方式采用如图 5.4 所示实数编码方式：个体 DNA（ $0.5 * N$ 个控制点）由一个 $N \times 1$ 的向量表示，前 $0.5 * N$ 个值代表横坐标，后 $0.5 * N$ 个代表纵坐标。一般 N 取 20 以上的偶数，曲线越复杂可设置的越高。

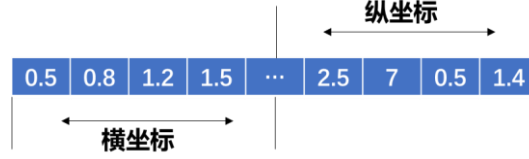


图 5.4 DNA 个体实数编码图

步骤 3：个体评价

本步骤是建立目标函数并计算种群中每个个体在目标函数下的适应度值。融合知识的目的是找到距离所有知识最短的知识曲线，本方法结合工程知识检验技术中的豪斯多夫距离被用于定义知识之间的相似程度，将适应度值设置为个体 DNA 拟合的贝塞尔曲线与所有知识曲线的豪斯多夫距离均值，如公式（5.1）所示：

$$Fitness\ function = \frac{1}{m} \sum_{i=1}^m H(Bezier(DNA), Exp_i) \quad (5.1)$$

m 为筛选后的知识的数量， DNA 指的遗传个体表示的控制点坐标， $Bezier(DNA)$ 指经过贝塞尔拟合公式（公式 3.4）得到的拟合曲线， H 是计算拟合曲线到某条知识曲线豪斯多夫距离。

步骤 4 个体选择

通过个体评价，确定种群个体的适应度值。本文以均匀排序的方法进行个体选择：按照适应度值由小到大排序，最终将前 200 个体作为下一代种群。

步骤 5 遗传操作

设定交叉概率 $p = 0.7$ 和变异概率 $q = 0.3$ ，确定交叉策略和变异策略。复制整个种群作为遗传操作对象，确定遗传操作的父代个体，以概率 p 完成交叉操作，以概率 q 完成变异操作，生成子代个体。

- 交叉操作：种群中每次都有种群规模为 $pop * p$ 个个体以确定好的交叉策略进行交叉操作。交叉操作的父代为整个种群的复制种群。
- 变异操作：种群中每次都有种群规模为 $pop * q$ 个个体进行以确定好的变异策略变异操作。变异操作的父代为整个种群的复制种群。

步骤 6 终止条件判断

种群进化 2000 代后终止。最终选择种群中适应度值最小的个体 DNA 所拟合曲线为最终的融合知识曲线。

5.3 基于有限点集费马点的点信息融合方法

经过第三章的贝塞尔曲线获取的单调与形状型工程知识曲线本身由一系列拟合的点构成，各个点的坐标信息是知识决定曲线形状的重要部分。不同于遗传算法从曲线整体去分析，基于有限点集费马点的工程知识融合方法则是从坐标点这一层面进行研究。首先将点之间的欧式距离表征点之间近似程度，距离越小表明坐标点越接近，构成的曲线相似程度更高。紧接着将各知识曲线统一划分成若干组切片，通过各切片的点集利用有限点集费马点的方法形成融合曲线点，最终由多个点形成最终的融合曲线。首先论述基于有限点集费马点的融合方法。

5.3.1 有限点集费马点定义

费马点问题(Fermat problem)是著名的几何极值问题，如图 5.5 所示它所描述的是：已知一个三角形，求作一点，使该点与这个三角形的三个顶点的距离之和最小。费马点即到三个点距离之和最小的点，推而广之，求到有限点距离之和最短的点可以定义成有限点集费马点问题。

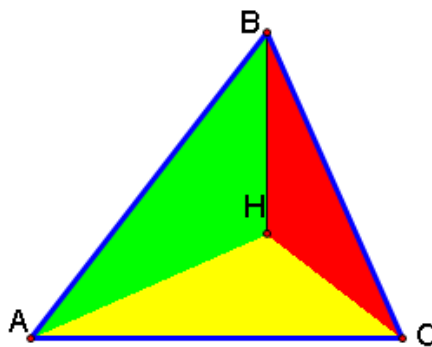


图 5.5 费马点问题

基于有限点集的费马点的融合方法意义在于，它使得融合点到所有的知识上对应位置的点的整体距离最短，使得该点的坐标信息尽可能的与所有点坐标相接近，从而保证了融合点构成的知识曲线与所有的曲线尽可能接近。

5.3.2 基于有限点集费马点的点信息融合方法

本方法的核心是将知识上的点划分成若干点集，利用求解有限点集费马点的方法求得各点集的融合点坐标，最后通过多个融合点构成最终的融合知识曲线。方法的详细步骤如下：

步骤 1：知识切片

结合公式 3.4，基于贝塞尔曲线获取方法使得单调与形状型知识曲线上点的数量由参数 t 决定，在此基础上统一知识获取流程中参数 t ，便能够保证两类知识曲线的点的数量一致。

因此针对以上两类知识的切片，按照 t 的取值进行划分：即每次 t 取值时，将两类知识拟合出的点划分为一个切片点集中，最终可形成由 t 采样数量决定的切片组 $[Points^1, Points^2, \dots, Points^t]$ ，其中每个点集包含 m 坐标点。

步骤 2：有限点集费马点求解

该步骤的核心是针对知识点集 $Points^t$ 求解其费马点 $Point_i$ 。目前针对有限点集费马点问题的解法有很多包括遗传算法，数值计算，几何图形解法等，本文采用“数值解法+迭代计算”的方法进行，主要方法如下：

(1) 定义问题函数：

针对费马点 $Point_i$ 定义距离之和最短的函数 $f(x, y)$ ，表达式为：

$$f(x, y) = \min \left(\sqrt{(x - x_i)^2 + (y - y_i)^2} \right) \quad (5.2)$$

其中 x_i, y_i 为点集 $Points^t$ 中第 j 个点的横纵坐标。

(2) 求导计算：

可以发现 $f(x, y)$ 是一个凸函数，其一阶导数的零点即为函数的最小值，因此问题转换为找到 $Point_j(x, y)$ 使 $f'(x, y)=0$ ，得到如 5.3 所示关系式（ y 与 x 表达式一致，仅是 y 与 x 的区别）：

$$\frac{\partial f(x, y)}{\partial x} = \sum_{i=1}^n \frac{x - x_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}} = 0 \quad (5.3)$$

(3) 迭代计算：

此时引入一个函数 $g(x, y) = f'(x, y) + (x, y)$ ，因此将问题进一步转换为： $(x, y) = g(x, y)$ ，通过不断的迭代计算使得 $g(g(x, y)) = (x, y)$ ，此时有限点集费马点坐标表达式转换为：

$$x = \frac{\sum_{i=1}^n \frac{x_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}}{\sum_{i=1}^n \frac{1}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}} \quad (5.4)$$

$$y = \frac{\sum_{i=1}^n \frac{y_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}}{\sum_{i=1}^n \frac{1}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}} \quad (5.5)$$

步骤 3：形成融合知识曲线

最终由 t 个费马点 \mathbf{Point}_i 构成融合知识曲线。

5.4 基于高斯过程的多条目工程知识融合

经过融合知识曲线尽管显著降低了知识的数目，其本身仍然存在一定不确定性，工程知识的共识信息并不能以融合曲线准确定义。这使得潜在的准确知识曲线大概率在融合知识附近。如何将不确定性信息进行量化表征是核心问题。本文利用高斯过程具有天然的不确定性量化表达能力对融合的知识曲线统一进行建模，利用概率模型表征知识信息。

步骤的核心是将融合知识曲线的横纵坐标分别作为输入，输出进行建模，得到最终的知识曲线概率表征模型，从而实现对于知识输入的快速预测。同时，知识之间的具有的差异性表现在曲线上各个点的位置不同，为充分考虑这种差异性，这里提取融合知识曲线上各点到所有知识对应位置点的最大差值作为高斯过程建模的噪声，如式（5.7）所示：

$$\varepsilon_i = 0.5 * (\max(\|\mathbf{Point}_i - \mathbf{point}_j^i\|) + \min(\|\mathbf{Point}_i - \mathbf{point}_j^i\|)) \quad (5.6)$$

同时为避免过大误差造成高斯过程模型预测的不确定性过大， ε_i 为融合知识曲线上各点与所有知识曲线上点的最大距离与最小距离的均值。最后参考式 4.6、式 4.7 中高斯过程的基本原理，将融合知识曲线的点坐标信息进行建模，并将 ε_j 作为额外噪声项 $\sigma^2 \mathbf{I}$ ，实现了最终融合知识的概率模型表征。

5.5 实验设计

5.5.1 实验目的

本章需要探究的实验目的概括如下：

（1）两种知识融合方法能否有效的保留多知识的有效信息，分析两种方法适用场景。

（2）围绕知识的筛选与融合方法，针对知识的处理能否有效提升融合进代理模型的预测精度，加快模型训练过程。

5.5.2 实验内容

5.5.2.1 实验案例

实验案例参考第四章的两个测试函数跟工程案例。具体介绍参考第四章测试函数与工程案例的介绍。

5.5.2.2 实验步骤

实验主要分为两部分进行，第一部分围绕着知识融合建模的方法能否有效保证知识的有效信息展开，第二部分的实验在前一实验基础上就知识融合进代理模型进行有效性验证。

➤ 实验 1

步骤 1：知识生成，参考第四章知识生成的相关步骤。

步骤 2：知识筛选，选择在基于梯度信息的知识检验方法对知识进行筛选。

步骤 3：知识融合，依据第 4 节基于遗传算法、基于有限点集费马点两种知识融合的方式，对步骤 2 中三种筛选方案得到的知识进行融合与建模。

步骤 4：预测检验，将各案例基准曲线的横坐标作为输入进行预测，并与基准线的纵坐标进行误差检验，由于高斯过程的预测包括均值和方差，将公式(4.20)的检验指标 $RMSE_p$ 作为预测误差。

步骤 5：重复步骤 2-4，完成各案例 20 轮实验，计算 20 组预测误差的平均值作为每种噪声值实验的最终误差，并结合 95%置信度水平，分析高斯过程模型能否有效表征基准知识曲线的有效信息。

➤ 实验 2

步骤 1-2：知识生成与筛选，实验 2 在基于实验 1 的实验结果上进行，首先选择三个案例 20 轮中误差最小的组为实验样本组，利用基于梯度信息的知识一致性度量指标进行知识的筛选，最终选择筛选出的前 50%的知识曲线为最终融合进代理模型的知识曲线。

步骤 3：知识融合，选择基于有限点集费马点的知识融合方法进行知识融合，基于高斯过程进行概率模型构建。

步骤 4：融合知识构建代理模型的有效性验证，选择将步骤 3 中表征融合知识的高斯过程模型融合进案例代理模型构建中，技术参考课题组叶文斌博士相关论文提出的基于进化神经网络的融合工程知识代理模型构建技术^[45]。

各案例中用于融合代理模型构建的数据样本均为随机采样 50 组数据。实验对照组为将步骤 2 筛选的知识曲线分别融合进代理模型中，以比较基于高斯过程的知识融合表征方法的优势。

5.6 实验结果与分析

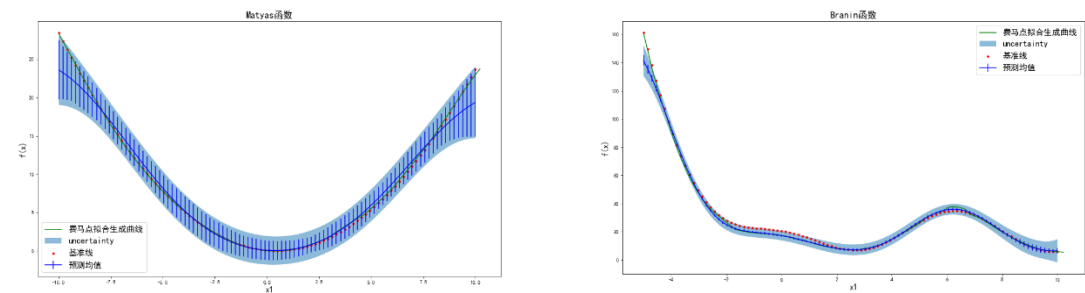
➤ 实验 1

如表 5.1 所示为车架案例、Matyas 测试函数、Branin 测试函数知识融合模型的预测误差值。由表可以发现遗传算法针对 Branin 测试函数时，预测误差相较其他明显更大，因而该方法不适用于这类曲线复杂同时坐标覆盖范围大的情形；但在曲线的坐标覆盖范围较小时有明显优势（车架案例归一化处理）。相较而言，基于有限点集费马点的融合方法在曲线坐标覆盖范围广时，各种曲线形状复杂程度下的误差表现均更低。

表 5.1 各案例基于两种方法融合预测误差

预测误差	遗传算法	有限点集费马点	有限点集费马点提升 (%)
车架	0.0198	0.0262	-24.4
Matyas 函数	1.729	1.414	22.3
Branin 函数	11.936	3.313	260.3

如图 5.6 所示依次为车架案例、Matyas 函数、Branin 函数在 20 轮中，最大预测误差组中，选择有限点集费马点作为知识融合方法的最终高斯过程的基准线预测图，可以发现建模在 95%置信度情况下，高斯过程预测的不确定性范围能有效覆盖基准线，保证了基准线的有效信息。



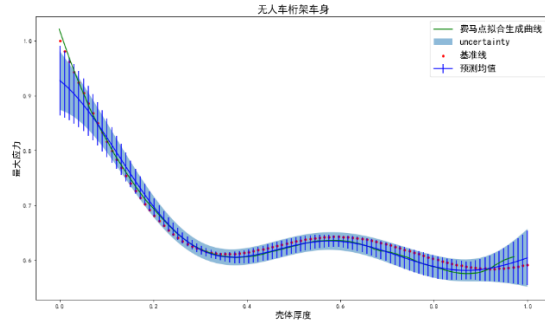


图 5.6 基于有限点集费马点融合方法的高斯过程基准线预测图

➤ 实验 2

图 5.7 展示的模型的训练误差图，其中“N-K”表示的为融合进工程知识的训练误差、“E-K”指的是将筛选完的工程知识分别融合进代理模型中的训练误差、“G-K”指的是先将工程知识融合成高斯过程后融合进代理模型中的训练误差。由图可以发现，融合工程知识能明显提升模型的预测精度，同时基于高斯过程融合工程知识的方法对知识进行融合处理后，代理模型的收敛速率更快，精度也有提升。

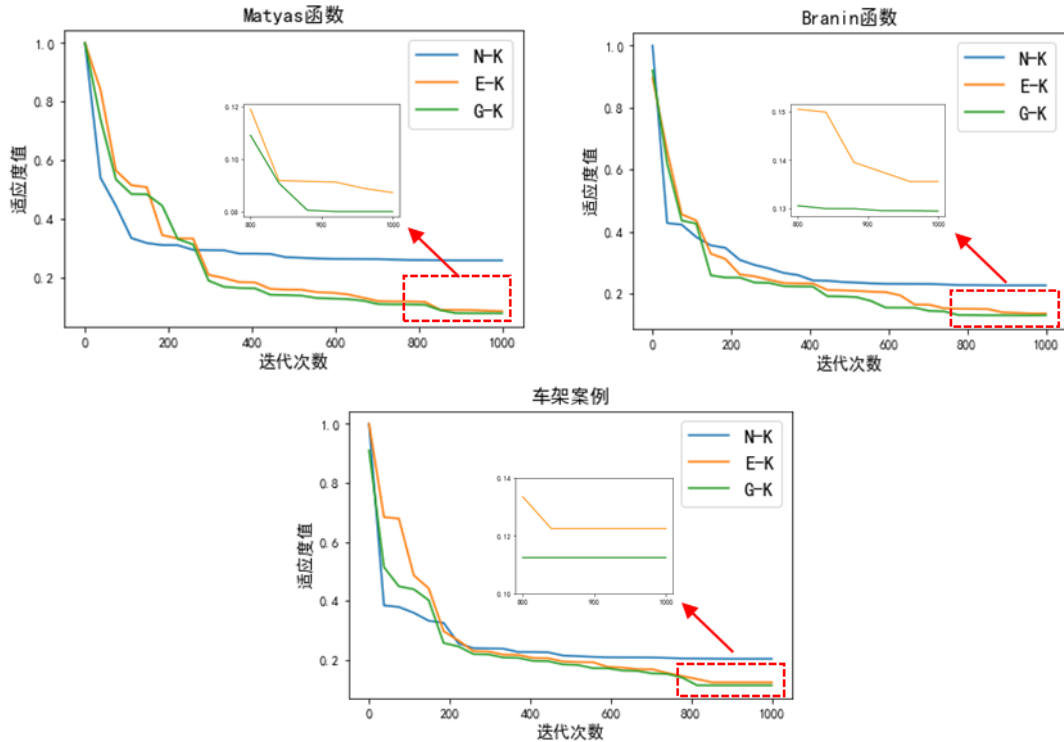


图 5.7 各案例代理模型训练误差图

5.7 本章小结

本章节要围绕基于筛选后的多条目知识的融合展开研究，首先提出了基于遗

传算法、有限点集费马点的工程知识融合方法,显著降低了知识的数量。接下来,结合高斯过程模型对融合知识曲线进行建模,实现多知识统一的概率模型表征。最后以测试函数和工程案例为例,设计实验来证明遗传算法的知识融合方法适合于形状简单的知识,对于形状复杂的知识效果不如有限点集费马点的方法。

结论

总结

本文针对融合工程知识构建代理模型以解决复杂装备设计问题中仿真周期长、仿真数据采集成本高等问题。面向专家围绕工程知识的获取、检验与融合为目标,开展交互式工程知识获取技术、基于多重一致性计算的工程知识检验技术和基于高斯过程的工程知识的融合技术等研究。本文的主要成果和结论概括如下:

(1) **实现基于贝塞尔曲线的交互式工程知识获取技术**,针对复杂工程设计问题梳理出属性型、单调型、形状型和公式型四种基本类型的知识,在此基础上利用文本匹配实现由知识文本到结构化知识模板的快速转换。进一步结合交互式的知识获取方法,专家可自主实现高效协同地知识获取,从而有效解决工程设计问题中知识来源广、参与专家人数众多导致的知识难以统一获取的难题。

(2) **提出基于多重一致性计算的工程知识检验技术**,在获取的不同类型的工程知识基础上,针对更具专家认知主观性的形状与单调类型知识,提出基于梯度信息、基于豪斯多夫距离两种知识度量指标,并结合 3Sigma 原则对异常的知识进行检验筛选。通过实验验证所提出方法的有效性,实验结果表明,两种方法均能有效筛选出异常的知识以提升知识的准确性。同时基于梯度信息的工程知识一致性检验方法更适用于构建代理模型的数据准确的情形,高精度的数据其反映变量之间的梯度信息能够有效地评估知识优劣。相反,在低精度的数据情况下,梯度信息难以保证,此时基于豪斯多夫距离的工程知识一致性检验方法更具有优势。

(3) **基于高斯过程的多条目工程知识融合技术**,为避免工程知识数量多造成最终融合代理模型过程中难以收敛的问题,本文所别提出的面向曲线信息的基于遗传算法的知识融合方法、面向点坐标信息的基于有限点集费马点的工程知识融合方法均能实现知识数量的降低,结合高斯过程模型的不确定性表征能力对融合知识进行建模以实现知识统一的概率模型表征,实现知识的差异性的概率表征。通过测试函数和工程案例进行验证,实验结果表明两种融合方法均能在降低知识数量的基础上表征知识的共识信息。由于遗传算法方法中个体 DNA 是对点坐标的编码,当形状型的知识曲线复杂时,个体 DNA 编码越复杂,遗传算法的计算成本更高,此时计算效率更高的费马点方法可优先考虑。

创新点

(1) **提出可交互的工程知识获取技术。**本文针对工程知识的特点,梳理出工程知识核心内容,针对其分类、表征等远未形成一致的理解与表征的问题,实现交互式的工程知识获取技术,面向专家实现快速的知识协同获取。

(2) **针对带有强烈专家主观认知的形状、单调类型知识提出检验和融合方法。**以上两种知识由于不具备准确的变量关系的描述,并不能像公式类型直接融合到模型中。本文针对知识特点,提出结合数据精度情形的两种知识度量方法,进一步利用 3Sigma 原则完成异常知识的筛选。同时为减少知识的数量,避免多知识的融合导致代理模型难以收敛的问题,提出面向曲线信息、点信息两种知识融合方法。最后基于高斯过程实现知识统一的概率模型表征,为融合工程知识构建代理模型提供用于应对多知识问题的处理方法。

展望

本文的研究工作围绕着工程知识的获取以及处理展开,研究专家获取的工程知识开始到知识的检验以及融合建模整个技术流程,并进行实验验证。研究的成果能够很好的应对融合工程知识构建代理模型中的涉及到知识的处理。但是在以下几个方面还需要进一步的研究:

(1) 多维输入的形状、单调型知识的修改方法

本文处理的形状与单调类型的知识主要是单输入与单输出之间的关系,形状涉及到二维空间。当输入涉及到多维时,如何有效的修改空间层面的曲线需要进一步的思考技术方法。

(2) 知识检验过程中的知识权重的研究

本文针对知识的检验过程中,对于知识本身采用相同的重要性程度。然而在实际的工程问题中,专家由于参与程度的不同对于知识本身有着不同的认知水平。因此可以考虑各专家对知识的贡献水平,引入专家认知权重系数,在度量步骤之前利用层次分析法进行知识的重要性加权研究。

(3) 面向知识度量方法的不同指标

本文在定义曲线一致性的指标过程中选择豪斯多夫距离指标,然而弗雷歇距离也作为一种常用于定义轨迹相似度的指标,可以考虑引入多种度量指标进行综合分析。

参考文献

- [1] 龙腾, 刘建, WANG G G.基于计算试验设计与代理模型的飞行器近似优化策略探讨[J].机械工程学报, 2016, 52(14): 79-105.
- [2] Ögren J, Gohil C, Schulte D. Surrogate modeling of the CLIC final-focus system using artificial neural networks[J]. Journal of Instrumentation, 2021, 16(05): P05012.
- [3] Smith R D, Chief Scientist M B. Simulation: The Engine Behind The Virtual World[J]. GEN, 1999, 1: 72.
- [4] 李伯虎, 柴旭东, 朱文海.复杂产品集成制造系统技术[J].航空制造技术, 2002(12): 17-20.
- [5] Sahnoun M H, Bettayeb B, Bassetto S J, et al. Simulation-based optimization of sampling plans to reduce inspections while mastering the risk exposure in semiconductor manufacturing[J], 2016, 27(6): 1335-1349.
- [6] Liu B, Koziel S, Zhang Q. A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems[J]. Journal of Computational Science, 2016, 12: 28-37.
- [7] Gong W, Zhou A, Cai Z. A Multioperator Search Strategy Based on Cheap Surrogate Models for Evolutionary Optimization[J]. IEEE Transactions on Evolutionary Computation, 2015, 19(5): 746-758.
- [8] Mallipeddi R, Lee M. An evolving surrogate model-based differential evolution algorithm[J]. Applied Soft Computing, 2015, 34: 770-787.
- [9] Hao J, Zhou M, Wang G, et al. Design optimization by integrating limited simulation data and shape engineering knowledge with Bayesian optimization (BO-DK4DO)[J]. Journal of Intelligent Manufacturing, 2020, 31(8): 2049-2067.
- [10] Shan S, Wang G G J S, Optimization M. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions[J], 2010, 41(2): 219-241.
- [11] Wang W, Welch W J. Bayesian Optimization Using Monotonicity Information and Its Application in Machine Learning Hyperparameter[J]. ArXiv, 2018, abs/1802.03532.
- [12] Hao J, Ye W, Wang G, et al. Evolutionary neural network-based method for constructing surrogate model with small scattered dataset and monotonicity experience[C]//2018 5th

- International Conference on Soft Computing & Machine Intelligence (ISCMI). IEEE, 2018: 43-48.
- [13] Han H, Wu X, Liu Z. Data-Knowledge-Based Fuzzy Neural Network for Nonlinear System Identification[J]. IEEE Transactions on Fuzzy Systems, 2019, 28(9): 2209-2221.
- [14] Park J, Sandberg I W. Universal Approximation Using Radial-Basis-Function Networks[J]. Neural Computation, 1991, 3(2): 246-257.
- [15] Friedman J H. Multivariate Adaptive Regression Splines[J]. The Annals of Statistics, 1991, 19(1): 1-67.
- [16] Joachims T. Making large scale SVM learning practical[J]. Advances in Kernel Methods: Support Vector Machines, 1999: 41-56.
- [17] 李焦赞, 高正红. 多目标进化算法和代理模型技术在气动稳健优化设计中的应用[J]. 空气动力学学报, 2012, 30(01): 46-51.
- [18] 韩忠华, 张瑜, 许晨舟, 等. 基于代理模型的大型民机机翼气动优化设计[J]. 航空学报, 2019, 40(01): 155-170.
- [19] 于梦阁, 潘振宽, 蒋荣超, 等. 基于近似模型的高速列车头型多目标优化设计[J]. 机械工程学报, 2019, 55(24): 178-186.
- [20] Qian Z, Seepersad C, Joseph V R, et al. Building Surrogate Models Based on Detailed and Approximate Simulations[J]. Journal of Mechanical Design - J MECH DESIGN, 2006, 128.
- [21] Fadel G M, Riley M F, Barthelemy J M. Two point exponential approximation method for structural optimization[J]. Structural optimization, 1990, 2(2): 117-124.
- [22] Jekel C F, Haftka R T. Testing Surrogate-Based Optimization with the Fortified Branin-Hoo Extended to Four Dimensions[J]. arXiv preprint arXiv:2107.08035, 2021.
- [23] Du X, Leifsson L. Optimum aerodynamic shape design under uncertainty by utility theory and metamodeling[J]. Aerospace Science and Technology, 2019, 95: 105464.
- [24] Koziel S, Leifsson L. Knowledge-based airfoil shape optimization using space mapping[C]//30th AIAA applied aerodynamics conference. 2012: 3016 .
- [25] Bennett A, Martin N, Avramova M. A surrogate model based on sparse grid interpolation for boiling water reactor subchannel void distribution[J]. Annals of Nuclear Energy, 2019, 131: 51-62.
- [26] Laskey M, Lee J, Fox R, et al. Dart: Noise injection for robust imitation learning[C]//Conference on robot learning. PMLR, 2017: 143-156 .
- [27] Longadge R, Dongre S J A. Class Imbalance Problem in Data Mining Review[J]. International

Journal of Computer Science & Network, 2013, 2(1): 23-29.

[28] 黄泽英, 李海艳, 林景亮. 迁移学习下的极限学习机代理建模方法及应用[J]. 计算机工程与应用, 2021, 57(22): 257-262.

[29] Guo Z, Tang J, Qiao J. An Improved Virtual Sample Generation Technology based on Mega Trend Diffusion[C]//2019 Chinese Automation Congress (CAC). IEEE, 2019: 2966-2972.

[30] Towell G G, Shavlik J W. Knowledge-based artificial neural networks[J]. Artificial Intelligence, 1994, 70(1): 119-165.

[31] Tversky A, Koehler D J. Support theory: A nonextensional representation of subjective probability[J]. Psychological Review, 1994, 101(4): 547-567.

[32] Lang B. Monotonic multi-layer perceptron networks as universal approximators[C]//International conference on artificial neural networks. Springer, Berlin, Heidelberg, 2005: 31-37.

[33] Abu-Mostafa Y S. Learning from hints in neural networks[J]. Journal of Complexity, 1990, 6(2): 192-198.

[34] Hinton G E. How neural networks learn from experience[J]. Scientific American, 1992, 267(3): 144-151.

[35] Sill J, Abu-Mostafa Y S. Monotonicity Hints[J]. Advances in Neural Information Processing Systems, 1997: 634-640.

[36] Sill J. Monotonic Networks[J]. Advances in neural information processing systems, 1997, 10.

[37] Rasmussen C, Williams C. Gaussian Process for Machine Learning[J], 2006.

[38] Daniels H, Velikova M. Monotone and Partially Monotone Neural Networks[J]. IEEE Transactions on Neural Networks, 2010, 21(6): 906-917.

[39] Riihimäki J, Vehtari A. Gaussian processes with monotonicity information[J]. Journal of Machine Learning Research - Proceedings Track, 2010, 9: 645-652.

[40] Gräve K, Stückler J, Behnke S. Learning Motion Skills from Expert Demonstrations and Own Experience using Gaussian Process Regression[M]. 2010: 1-8.

[41] Simsek M. Knowledge based three-step modeling strategy exploiting artificial neural network[M]//Solving Computationally Expensive Engineering Problems. Springer, Cham, 2014: 219-239.

[42] Lenk P J, Choi T. Bayesian analysis of shape-restricted functions using Gaussian process

- priors[J]. *Statistica Sinica*, 2017: 43-69.
- [43] Gupta M, Bahri D, Cotter A, et al. Diminishing returns shape constraints for interpretability and regularization[J]. *Advances in neural information processing systems*, 2018, 31.
- [44] Wu X, Han H, Liu Z, et al. Data-knowledge-based fuzzy neural network for nonlinear system identification[J]. *IEEE Transactions on Fuzzy Systems*, 2019, 28(9): 2209-2221.
- [45] Hao J, Ye W, Jia L, et al. Building surrogate models for engineering problems by integrating limited simulation data and monotonic engineering knowledge[J]. *Advanced Engineering Informatics*, 2021, 5: 1-47.
- [46] Yu L, Wang L, Yu J. Identification of product definition patterns in mass customization using a learning-based hybrid approach[J]. *The International Journal of Advanced Manufacturing Technology*, 2008, 38(11): 1061-1074.
- [47] Armaghan N, Renaud J. An application of multi-criteria decision aids models for case-based reasoning[J]. *Information Sciences*, 2012, 210: 55-66.
- [48] Juarez J M, Riestra T, Campos M, et al. Medical knowledge management for specific hospital departments[J]. *Expert Systems with Applications*, 2009, 36(10): 12214-12224.
- [49] Azadeh A, Ebrahimipour V, Bavar P. A fuzzy inference system for pump failure diagnosis to improve maintenance process: The case of a petrochemical industry[J]. *Expert Systems with Applications*, 2010, 37(1): 627-639.
- [50] D'eredita M A, Barreto C. How does tacit knowledge proliferate? An episode-based perspective[J]. *Organization Studies*, 2006, 27(12): 1821-1841.
- [51] Liu L, Jiang Z, Song B. A novel two-stage method for acquiring engineering-oriented empirical tacit knowledge[J]. *International Journal of Production Research*, 2014, 52(20): 5997-6018.
- [52] Chan F T. Application of a hybrid case-based reasoning approach in electroplating industry[J]. *Expert Systems with Applications*, 2005, 29(1): 121-130.
- [53] Argote L, Miron-Spektor E. Organizational learning: From experience to knowledge[J]. *Organization science*, 2011, 22(5): 1123-1137.
- [54] 张锡爱, 马翥. 主动知识辅助系统中制动蹄设计知识库的构建[J]. *计算机工程与科学*, 2013, 35(1): 175-179.
- [55] 王定钦, 刘建军, 黎展滔. 基于模具经验性知识的本体编辑器系统设计与实现[J]. *机电工程技术*, 2011, 40(12): 65-69.

- [56] 袁晓征, 毛宁, 陈庆新, 等. 一类经验性知识解释器的设计与实现[J]. 计算机技术与发展, 2009, 19(10): 242-246.
- [57] 韦于莉. 知识获取研究[J]. 情报杂志, 2004(04): 41-43.
- [58] Castro-Schez J J, Murillo J M, Miguel R, et al. Knowledge acquisition based on learning of maximal structure fuzzy rules[J]. Knowledge-Based Systems, 2013, 44: 112-120.
- [59] Liu W-Y, Yue K, Gao M-H. Constructing probabilistic graphical model from predicate formulas for fusing logical and probabilistic knowledge[J]. Information Sciences, 2011, 181(18): 3828-3845.
- [60] Feng L, Li T, Ruan D, et al. A vague-rough set approach for uncertain knowledge acquisition[J]. Knowledge-Based Systems, 2011, 24(6): 837-843.
- [61] Do Rosário C R, Kipper L M, Frozza R, et al. Modeling of tacit knowledge in industry: Simulations on the variables of industrial processes[J]. Expert Systems with Applications, 2015, 42(3): 1613-1625.
- [62] Cairó O, Guardati S. The KAMET II methodology: Knowledge acquisition, knowledge modeling and knowledge generation[J]. Expert Systems with Applications, 2012, 39(9): 8108-8114.
- [63] Jin Y, Ishino Y. DAKA: design activity knowledge acquisition through data-mining[J]. International Journal of Production Research, 2006, 44(14): 2813-2837.
- [64] 韩爱玲, 宁献文, 张田甜, 等. 基于专家系统理念的航天器特殊控温策略推理测试验证[J]. 航天器环境工程, 2021, 38(05): 569-574.
- [65] Ruiz P P, Foguem B K, Grabot B. Generating knowledge in maintenance from Experience Feedback[J]. Knowledge-Based Systems, 2014, 68: 4-20.
- [66] 马艳, 代志龙, 危辉. 基于场景及规则获取模板的知识获取方法[J]. 计算机工程与设计, 2010, 31(09): 1995-1997.
- [67] Cheah W P, Kim Y S, Kim K-Y, et al. Systematic causal knowledge acquisition using FCM constructor for product design decision support[J]. Expert Systems with Applications, 2011, 38(12): 15316-15331.
- [68] Wang J, Wu Y, Liu X, et al. Knowledge acquisition method from domain text based on theme logic model and artificial neural network[J]. Expert Systems with Applications, 2010, 37(1): 267-275.
- [69] 王平, 张勤. 基于模糊聚类的电力专家故障诊断信息融合方法[J]. 黑龙江电力, 2020, 42(02): 109-112.

- [70] Kwong E, Lee W B. Knowledge elicitation in reliability management in the airline industry[J]. *Journal of Knowledge Management*, 2009.
- [71] 尹光辉. 基于神经网络的装甲车辆维修系统[J]. *车辆与动力技术*, 2015(02): 29-31.
- [72] 陈洪雁, 张大伟, 杨欢, 等. 航天军事专家智能知识服务系统的设计与实现[J]. *现代电子技术*, 2021, 44(19): 8-12.
- [73] 刘兆政, 肖明清, 胡阳光, 等. 基于流向图的不完备信息下导弹故障诊断知识获取[J]. *计算机测量与控制*, 2021, 29(01): 126-129.
- [74] 张英, 王儒敬, 魏圆圆. 基于知识地图的隐性知识获取系统的设计[J]. *微电子学与计算机*, 2009(5): 17-20.
- [75] Rucklidge W J. Efficiently Locating Objects Using the Hausdorff Distance[J]. *International Journal of Computer Vision*, 1997, 24(3): 251-270.
- [76] 聂斌, 孙会东, 李佩, 等. 基于改进豪斯多夫距离的非参数轮廓变点识别[J]. *中国机械工程*, 2015, 26(08): 1029-1034.
- [77] 余有明, 刘玉树, 阎光伟. 遗传算法的编码理论与应用 [J]. *计算机工程与应用*, 2006(03): 86-89.
- [78] 李旺红. 异构网络中基于机器学习的网络选择算法[D]. 南京邮电大学, 2019.

攻读学位期间发表论文与研究成果清单

➤ 参与项目

国家重点研发计划“工程知识与数据融合驱动的复杂产品一体化智能设计方法”
(2021YFB1714500)

GF973 科研项目“基于自学习 xxx 气动外形气动预测技术”
(2019-JCJQ-ZD-049-02)

➤ 发表论文

“Bezier Curve-based shape knowledge acquisition and fusion for surrogate model construction” 国际会议论文 (FTC2022), 已录。

致谢

白驹过隙，时光荏苒。转眼间，在北京理工大学三年的悠悠岁月即将在这个鸟语蝉鸣的季节画上一个句号，而这于我的人生，只是一个逗号。我将满载着北理工给予的学识和学风，老师们教导的严谨和谦逊，怀揣感恩之心，踏上新的征程。犹记得三年前，迎着初秋宜人的清风，伴着依然稚嫩的笑脸，怀着对于学术的敬畏，带着对于未来的憧憬，体会着梦飞的时刻，我踏进了北理工这所向往已久的双一流大学。此时此刻，论文完成在即，心中感慨颇多，最想要表达的就是感谢，表达对所有在我研究生期间给予帮助的老师、同学、家人和朋友们衷心的感谢！

首先要感谢我的导师樊红丽老师、郝佳老师，老师们一直治学严谨，平易近人，在我研究生生涯中给予了我极大的支持和帮助，在老师们的悉心指导下，我不断进步，对学术和科研的了解不断加深，从一个懵懂的本科生成长为一名积极的研究生。在毕业论文的写作过程中，两位老师总是不厌其烦地为我解答疑惑，耐心修改我的论文并提出中肯的意见。在学术上耐心的点拨和鞭策，在生活上细心的关心和爱护，让在异乡求学的我收获匪浅、倍感温暖，在此谨向两位老师表达我由衷的感谢以及深深的敬意！感谢课题组的王国新老师，为我们营造了良好的生活环境和积极向上的学习氛围。一直以来，王老师为课题组注入源源不断的活力，为我们的科研提供强有力的支持。感谢卢继平老师一直以来对我在学习上的帮助，针对我的研究困惑一直给予方向上的指导，认真耐心的解决我的困惑。

感谢北理工嘉兴长三院的孙华飞老师、袁旭老师、孙渊老师在嘉兴的学习生活中对我找工作、研究学习的指导与帮助。感谢工业与智能系统工程研究所的阎艳老师、姚丽亚老师、明振军老师、胡耀光老师、薛庆老师、宫琳老师、敬石开老师、张发平老师、段晓峰老师、闻敬谦老师、刘敏霞老师，诸位老师组成了工业与智能系统工程这个大家庭，共同为同学们提供良好的科研环境，给予我们研究工作和论文撰写上的精心指导以及热情鼓励。感谢王儒师兄、黄思翰师兄，两位博士后师兄对每周的组会内容悉心策划，并对我的科研方向提供了许多帮助和建议。感谢贾良跃师兄、叶文斌师兄带领着智能设计小组奋发前进、全面发展，督促我们做好科研的同时，也积极参加各项学科竞赛，取得优异的成绩。

感谢实验室所有的师兄师姐、师弟师妹们以及同窗好友们，感谢孙嘉伟师兄、牛红伟师兄、龙辉师兄在项目工作上的帮助和指导；感谢已经毕业的霍阳阳师兄、

周梦莹师姐、曾梦梦师姐、郭江敏师兄、余士严师兄一致鼓励与帮助。感谢杨思宇、吉庆、朱志成、王嘉蕙、王子墨、赵雷同学一直以来的包容和支持。感谢蒲雪宁、张馥琳、黄晋等学弟学妹们在科研、项目和生活上的理解和帮助。

感谢呵护我成长的父母，当我一路跌跌撞撞前行的时候，是你们一直在我身后给予无私的支持和鼓励，坚定了我不断奋斗的信念！

同时，本文的完成，参考了许多前辈和同行们的研究成果，向引用文献的作者及献身科研教育的学者致敬！

最后，衷心感谢各位专家教授在百忙之中抽出时间评阅本文，并提出宝贵意见。

二零二二年五月
于北京理工大学长三角研究院