

分类号：TM74

单位代码：10363

密类级：公开

学位代号：2220320169

题 目

多维工艺信息融合的典型

网状工艺路线发现方法研究

题 目

Research on Typical Network Process

Route Discovery Method Based on Multi-

Dimensional Process Information Fusion

论文作者：

王梦娇

指导教师：

徐彬梓

专 业：

电子信息

研究方向：

控制工程

论文答辩日期：

2025 年 5 月 7 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：王梦妍

日期：2025年5月14日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 _____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：

王梦妍

指导教师签名：

徐林祥

日期： 2025 年 5 月 14 日

日期： 2025 年 5 月 14 日

多维工艺信息融合的典型网状工艺路线发现方法研究

摘 要

工艺规划过程的好坏直接决定了后续实际生产过程的质量，而传统基于技术人员自身经验的工艺规划过程过于严重依赖其自身能力，并不利于知识的重复利用。为此，在工业大数据和知识自动化的发展下，基于计算机辅助工艺规划(Computer Aided Process Planning, CAPP)的工艺规划方法受到了广泛的关注。在派生式工艺规划过程中，典型工艺路线的有效发现可以为 CAPP 的检索过程提供更加精准且充足的工艺信息，从而提高后续工艺规划的质量。但是现有相关研究还存在两个问题：一是有关线性工艺路线的理论方法由于没有考虑网状工艺路线的结构复杂性，既无法直接应用，也难以合理量化网状工艺路线中的各类工艺信息。二是大部分现有研究忽略了典型工艺路线发现中的聚类有效性问题，对于这一更加复杂的聚类问题缺乏有效的算法设计。鉴于此，本文提出了一种基于多维工艺信息融合的典型网状工艺路线自动化发现方法，其具体内容如下：

1. 针对网状工艺路线的相似度度量问题，在信息需求分析的基础上，为网状工艺路线中蕴含的四种工艺信息设计了不同的量化方法，并通过主成分分析(Principal Component Analysis, PCA)将其合成综合相似度。
2. 基于上述相似度度量的构建，并考虑聚类有效性，在传统近邻传播(Affinity Propagation Clustering Algorithm, AP)算法中引入了火鹰

优化算法(Fire Hawk Optimizer, FHO), 以优化 AP 算法的参考度与阻尼系数, 从而在聚类结果与软约束之间取得平衡, 以发现更符合实际需求的典型网状工艺路线。

3. 仿真实验表明, 本文所提网状工艺路线的相似度度量能有效区分各种相似度情况, 具有更高的灵敏性。同时, 引入的 FHO 确实能提高 AP 算法的聚类效果, 其中 FHO-IAP 在两种数据集上都表现出更好的聚类效果, 具有较强的稳健性。

关键词: 典型工艺路线, 相似度度量, CAPP, AP 算法, FHO

RESEARCH ON TYPICAL NETWORK PROCESS ROUTE DISCOVERY METHOD BASED ON MULTI- DIMENSIONAL PROCESS INFORMATION FUSION

ABSTRACT

The quality of process planning directly determines the quality of the subsequent actual production process, while the traditional process planning process based on the experience of technicians relies too heavily on their own ability, and is not conducive to the reuse of knowledge. Therefore, with the development of industrial big data and knowledge automation, the Process Planning based on Computer Aided Process Planning (CAPP) has received extensive attention. In the process of derived process planning, the effective discovery of typical process routes can provide more accurate and sufficient process information for the retrieval process of CAPP, thus improving the quality of subsequent process planning. However, there are still two problems in the existing research. First, because the theoretical method of linear process route does not consider the structural complexity of the network process route, it cannot be directly applied, and it is difficult to quantify the various process information in the network process route. Second, most of the existing studies ignore the problem of clustering efficiency in typical process route discovery, and lack effective algorithm design for this more complex

clustering problem. In view of this, this paper proposes a typical network process route automatic discovery method based on multi-dimensional process information fusion, which is detailed as follows:

1. Aiming at the similarity measurement problem of the network process route, on the basis of information demand analysis, different quantification methods were designed for the four kinds of process information contained in the network process route, and the comprehensive similarity was synthesized through Principal Component Analysis (PCA).

2. Based on the construction of the above similarity measure and considering the clustering effectiveness, the Fire Hawk Optimizer (FHO) is introduced into the traditional Affinity Propagation algorithm (AP). The reference degree and damping coefficient of AP algorithm are optimized to strike a balance between clustering results and soft constraints, so as to find a typical network process route that is more in line with actual requirements.

3. Simulation experiments show that the similarity measurement of the network process route proposed in this paper can effectively distinguish various similarity situations and has higher sensitivity. At the same time, the introduced FHO can indeed improve the clustering effect of AP algorithm, among which the FHO-IAP shows better clustering effect on both data sets, and has strong robustness.

KEY WORDS: typical process route, similarity measurement, CAPP, AP algorithm, FHO

目录

摘 要.....	I
ABSTRACT.....	III
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 研究现状分析.....	2
1.2.1 人工智能技术在工艺规划中的应用.....	2
1.2.2 工艺路线的相似度设计.....	3
1.2.3 工艺路线的聚类分析.....	5
1.3 本文主要研究内容.....	7
第 2 章 典型网状工艺路线发现问题的需求分析.....	9
2.1 引言.....	9
2.2 典型工艺路线发现的总体框架与思路.....	9
2.3 网状工艺路线相似度的问题难点.....	10
2.3.1 网状工艺路线相似度设计的信息需求分析.....	10
2.3.2 网状工艺路线相似度案例库构建.....	13
2.4 典型网状工艺路线发现的问题难点.....	15
2.4.1 典型网状工艺路线的聚类有效性需求分析.....	15
2.4.2 考虑聚类有效性的典型网状工艺路线发现问题数学描述.....	16
2.5 本章小结.....	18
第 3 章 多维工艺信息融合的网状工艺路线相似度设计.....	19
3.1 引言.....	19
3.2 常见工艺路线相似度度量.....	19
3.2.1 基于最长公共子序列算法的相似度.....	19
3.2.2 基于合规指数的相似度.....	20
3.2.3 多维工艺路线最长公共子序列检索算法.....	20
3.2.4 采用距离公式计算工艺路线相似度.....	22
3.3 多维工艺信息融合的新型网状工艺路线综合相似度度量.....	23
3.3.1 工序间相似度度量.....	24

3.3.2 先后顺序关系度量	24
3.3.3 共同工序数计算	26
3.3.4 结构复杂度计算	26
3.3.5 多维工艺信息融合的新型网状工艺路线综合相似度	27
3.3.6 相似度度量性能比较	29
3.4 本章小结	32
第 4 章 考虑聚类有效性的典型网状工艺路线发现.....	33
4.1 引言	33
4.2 所用算法的基础理论	33
4.2.1 FHO 算法	33
4.2.2 OAP 算法	36
4.2.3 IAP 算法	39
4.3 基于 FHO 优化 AP 的典型工艺路线发现方法	42
4.3.1 总体思路与框架	42
4.3.2 基于 FHO-OAP 的典型工艺路线发现方法	43
4.3.3 基于 FHO-IAP 的典型工艺路线发现方法	45
4.4 仿真实验验证	47
4.4.1 聚类性能指标	47
4.4.2 仿真实验设置	47
4.4.3 基于小型数据集的性能分析	48
4.4.4 基于大型数据集的性能分析	51
4.5 本章小结	54
第 5 章 总结与展望.....	55
5.1 总结	55
5.2 展望	55
参考文献.....	57
攻读硕士期间取得的学术成果.....	62
致谢.....	63
附录.....	64

第 1 章 绪论

1.1 研究背景及意义

随着经济的不断发展,制造业迫切需要一种敏捷且有效的工艺规划方法来快速响应不断变化的市场需求^[1],作为产品设计与实际生产间的枢纽,工艺路线的优劣直接决定了制造系统的生产质量与效率^[2,3]。计算机辅助工艺规划(Computer Aided Process Planning, CAPP)作为机械零件设计和制造之间的纽带,对产品质量、加工效率和最终生产成本有着直接而重要的影响^[4]。常见的基于 CAPP 的工艺规划方法包括三种:

(1) 在制造业的工艺规划领域,构建了一种基于检索式的标准工艺规划体系,该体系通过系统性编号与计算机存储,实现了对预设计零件标准工艺的有效归档。当制定零件工艺过程时,该体系能根据输入的零件信息,智能检索并匹配出最合适的标准工艺,从而极大地提升了工艺规划的效率与准确性,促进了制造业的标准化、智能化发展。

(2) 在先进制造技术的研究范畴内,派生式工艺规划过程借助零件之间存在的相似特性展开。其核心原理在于,具有相似性的零件往往对应着相似的工艺过程。基于此,在实际工艺规划时,通过对相似典型零件已有的工艺过程进行精准检索,随后依据目标新零件的具体特征与要求,对检索到的工艺过程进行针对性的增添、删减或细致编辑等操作,最终衍生出适用于该新零件的工艺过程。

(3) 创成式工艺规划过程区别于派生式系统,其独特之处在于,该过程依据输入的零件信息,借助系统所集成的工程数据以及特定的决策方法,自动完成零件工艺过程的生成。在这一过程中,系统并非像派生式系统那样依赖已有的相似典型零件工艺过程,而是凭借自身所具备的数据资源与决策逻辑,独立自主地规划出适合输入零件的工艺路线,展现出高度的智能化与自主性。

由此可见,对于基于 CAPP 的派生式工艺规划而言,检索的典型工艺路线质量直接影响其派生的新零件的质量,故如何更加有效的从海量工业数据中挖掘潜在的典型网状工艺路线,是本文的最终目的。

典型工艺路线本质上是零件簇的工艺模板,它包含了零件簇中所有历史工艺路线之间高度重合的共同工艺信息与知识^[5]。从集群的角度而言,典型工艺路线是零件簇的中心点,它可以有效总结和代表同一集群中其他工艺路线的内在信息。这种基于典型工艺路线的规划方法不仅可以有效提升新产品工艺规划的效率和质量,而且有利于工艺路线标准化。作为 CAPP 的基石,检索所得典型工艺路线的质量会直接影响后续工艺规划的效果^[6]。

在海量历史工艺路线数据中发现典型工艺路线主要包含两个步骤:相似度设计和聚类分析^[7]。设计合适的相似度计算方法可以敏锐地量化不同工艺路线之间的共有信息,为后续的聚类分析提供细粒度的相似度度量,从而保证所得典型工艺路线的质量。另一方面,聚类分析是将历史工艺路线按照其相似度进行分组的过程,以对海量工艺路线进行有效的归纳和总结并将各聚类中心提取作为典型工艺路线^[8]。在这一过程中,如果相似度设计或聚类的效果不好,可能会导致工艺路线分析结果不准确,从而得到不合理的典型工艺路线,无法有效地指导后续的工艺规划过程^[9]。

目前有关典型工艺路线的研究大多数都是针对线性工艺路线而展开的,忽视了网状工艺路线这一更符合实际且更加柔性的工艺表现形式^[10]。不同于线性工艺路线,网状工艺路线具有易于系统重新配置和减少转换时间等优点^[11],并且能够很好面对实际生产中遇到的各类突发状况^[12]。因此,研究典型网状工艺路线的自动化发现问题,更契合当前市场化、柔性化、智能化的生产需求^[13, 14]。

1.2 研究现状分析

1.2.1 人工智能技术在工艺规划中的应用

人工智能(Artificial Intelligence, AI)技术近年来在工艺规划领域展现出巨大潜力,极大地推动了传统制造向智能化方向的转型^[15]。工艺规划作为将产品设计要求转化为具体制造过程的重要环节,其核心任务包括工艺方案的制定、加工顺序的优化和资源的高效配置。然而,传统的工艺规划方法通常依赖于人工经验和数据库,面对复杂多变的现代制造环境,这些方法逐渐显现出效率低、灵活性差、不利于知识传承问题。

AI 技术的引入为工艺规划提供了全新的思路。基于人工智能的工艺规划方法能够利用历史工艺数据和领域知识,结合机器学习和深度学习算法,完成从数据挖掘到工艺方案生成的全过程^[16]。在此过程中, AI 技术可以通过历史数据的分析提取加工规律,进而优化工艺参数和提高生产效率^[17]。

Yu 和 Fan 等人^[18]结合了机器视觉检测技术和自动磨削技术,使用自动磨削和修复平台来检测和修复圆柱形零件的缺陷。针对工件初始安装偏心问题,提出了一种可调节砂轮进给量和磨削位置的磨削路径规划算法。Zhan 和 Wu 等人^[19]提出了一种将知识图谱与工艺规划中的工艺决策活动相结合的强化学习方法,挖掘工艺知识图谱的学习潜力,还创建了结合工艺规划环境的复合奖励函数。Wang 和 Zheng^[20]提出了一种新的具有工序柔性的能量感知再制造工艺规划与调度模型。该模型不仅同时解决了再制造过程的规划和调度问题,而且采用了机器速度切换策略来降低能耗。Cheng 和 Zhao 等人^[21]提出了一种结合 STEP-NC 和基于特征的智能工艺规划解决方案。通过将加工特征作为规划单元并开发集成的内置模型和算法来帮助决策,将实现更加自动化的操作。Guo 和 Liu 等人^[22]为了克服在工艺规划问题中寻找最优或接近最优解以及适应实际生产中动态资源变化的挑战,引入了一种用于工艺规划问题的多起始节点探索的深度强化学习方法。

在生成式工艺规划领域,典型工艺路线发现是一个关键环节。其核心目标是通过分析历史工艺数据,提取出具有代表性的工艺路线,作为新产品工艺规划的参考。这一方法不仅能够减少规划时间,还可以提高工艺方案的合理性和可行性。然而,典型工艺路线发现当前面临两个主要问题:(1) 是工艺路线相似度的度量问题,(2) 是基于相似度的工艺路线聚类问题。相似度计算是影响工艺路线匹配质量的关键,而有效的聚类方法则是提取典型工艺路线的基础。这两个问题的解决将直接影响生成式工艺规划的精度和效率,也为进一步推动工艺规划的智能化发展奠定了基础。

1.2.2 工艺路线的相似度设计

作为典型工艺路线发现的基础^[23],线性工艺路线的相似度已经有大量前期研究。这些研究在分析线性工艺路线的内在信息与特征基础上,提出了不同的度量方法来计算线性工艺路线直接的相似度。

Choobineh^[24]最先研究了工序间的先后顺序关系，他提出了一种用于设计单元制造系统的两阶段方法。第一阶段通过零件族划分来完成零件的分类。在该阶段，建议采用基于新的邻近度量的聚类技术。该邻近度量综合考虑了制造工序及其工序顺序，从而更精确地反映零件之间的加工相似性。第二阶段则进行设备单元划分，为此提出了一个整数规划模型。该模型的求解将明确每个单元中设备的类型和数量，并确定零件族与单元的分配关系。

Zhou 和 Dai^[25]提出了一种基于零件特征的粒计算与相似性算法的典型工艺路线发现方法。首先，通过分析和比较零件的加工特征及其特征间的拓扑关系，构建了工艺序列相似性计算模型，并据此建立了所有工艺序列的模糊相似矩阵。随后，根据粒计算理论中的模糊商空间模型，构造了具有层次结构的工艺路线商空间族。通过信息熵对每个工艺路线商空间的粒度进行度量，并计算相邻商空间之间的信息熵差值，即信息增益。最后，选择具有较大信息增益且粒度内工艺序列最小相似度较高的商空间层作为最优工艺信息粒层，并基于最长公共子序列算法（Longest Common Subsequence Algorithm, LCS）从该粒层中的工艺信息粒中提取典型工艺路线。

Liu 和 Zhang 等人^[26]采用聚类分析作为挖掘典型工艺路线的算法，并通过数据矩阵建立了描述工艺路线的数学模型。在工艺路线聚类过程中，涉及以下三种相似性计算：一是工序间相似性：基于工序编码使用曼哈顿距离（Manhattan Distance）进行计算。二是工艺路线间相似性：通过欧几里得距离计算，并以不相似矩阵表示。三是工艺路线簇间相似性：基于不相似矩阵的平均距离进行评估。但是他们都忽略了工序相似度和先后顺序关系的相关性，只使用线性加权和来得到最终的结果。

Xu 和 Wang 等人^[27]等人分析了典型工艺路线知识发现所需的必要工序信息，并以此提出了一种新的线性工艺路线综合相似度度量方法。他创造性地将工序间相似度与 LCS 相结合，提出伪 LCS 的概念，并通过 Kuhn-Munkres (KM) 算法对其进行了统一度量。

尽管线性工艺路线已经被深入研究，但这些为线性工艺路线设计的相似度度量不能直接用于网状工艺路线，因为网状工艺路线内部结构更为复杂且灵活，需

要考虑的工艺信息更多。此外，生产过程中的多种工艺信息与网状工艺路线的复杂结构相互耦合，想要有效度量这些工艺信息就显得更加困难。

据我们所知，Navaei 和 Elmaraghy^[28]最先研究了网状工艺路线的相似度度量问题，他们为网状工艺路线设计了一种新的相似度系数，在这其中考虑了先后顺序关系、共同工序数和体积相似性，并通过线性加权进行组合。但是，他们所提计算相似度的方法将所有工序都看成完全不同的情况，而忽视了工序间本身存在的相似关系。

1.2.3 工艺路线的聚类分析

聚类分析是一种重要的无监督学习方法，广泛应用于数据挖掘和模式发现领域。其目标是将数据对象划分为若干个簇，使得同一簇内的对象相似度较高，而不同簇之间的对象相似度较低。根据实现方式的不同，常见的聚类方法可以分为以下几类：

（1）基于划分的聚类方法。这类算法依赖预设的簇数量，通过优化簇内样本的紧密度实现分组，适合规则分布的数据。K-Medoids 就是一种常见的基于划分的算法^[29]，它通过迭代优化使得每个簇的质心和簇内数据点的平方和最小，并且使用实际的数据点作为中心点，减少噪声和异常值的影响。

（2）基于层次的聚类方法。基于层次的聚类方法包括凝聚层次聚类（Agglomerative Hierarchical Clustering, AHC）^[30]和分裂层次聚类（Divisive Hierarchical Clustering, DHC）^[31]两种，它们通过构建树状结构逐步揭示数据的多层次关系。凝聚层次聚类是从每个点作为一个簇开始，不断合并相似的簇。分裂层次聚类则是从一个整体簇开始，不断将其分裂为小簇。层次聚类的结果常常表示为树状结构，不同层次的聚类结果表示不同的聚类个数。

（3）基于密度的聚类方法。这类算法通过识别数据点密集区域来形成簇，可以有效处理噪声和形状复杂的簇。常见的算法有 DBSCAN（Density-Based Spatial Clustering of Applications with Noise）算法和 OPTICS（Ordering Points To Identify the Clustering Structure）算法。DBSCAN 算法通过密度连接形成簇，能够自动识别簇的数量，并处理噪声。OPTICS 算法扩展了 DBSCAN，可以发现不同密度的簇。

在工艺路线发现中,聚类算法主要作用是基于输入的工艺路线相似度矩阵对海量历史工艺路线进行聚类分析,从而发现其中隐含的共有模式。因此,设计好网状工艺路线的相似度度量后,对其进行聚类分析是发现典型工艺路线的第二个步骤。典型工艺路线发现的目的是通过海量历史工艺路线的有效聚类,从每个零件簇中挖掘和发现最具共性和代表性的那一条工艺路线,作为最后输出的典型工艺路线。因此,相较于传统聚类问题是将数据归类而言,典型工艺路线的发现问题是找出各聚类簇中包含最多共性工艺信息的那一条真实工艺路线。

目前为止,国内外对于工艺路线的聚类问题已有了十分深入的研究。Wu 和 Suzuki^[32]提出一种新的两阶段聚类方法,在第一阶段中,基于工序间相似性等不同准则对聚类簇进行识别;在第二阶段中,通过建立数学模型,将聚类簇分配到不同机器上,使总成本最小化,并采用不同的测试问题和敏感性分析并检验了其有效性。王琳和张永健等人^[33]等人通过近邻传播(Affinity Propagation Clustering Algorithm, AP)算法对工艺过程进行聚类分析,并根据 Silhouette 指标和 In-Group Proportion (GP)指标对聚类结果进行评价分析,得到有效聚类结果和最佳聚类数目。Li^[7]提出了一种基于多级信息熵的机械加工工艺路线相似度计算方法。在此基础上,他还提出了一种基于谱聚类和粒子群优化 K-means 算法的工艺路线聚类模型,根据所得相似度实现了对工艺路线的有效聚类。

然而,上述研究都忽视了典型工艺路线发现问题中聚类的有效性。Gupta 和 Seifoddini^[34]首先提出并深入研究了典型工艺路线发现问题的聚类有效性,但他们的研究采用了硬约束,使得在聚类过程中对约束条件及其参数的设置要求较高,无形中加剧了人为设定约束参数的难度。文献[35]针对这一情况在 AP 算法中引入了数量软约束和半径软约束来保证聚类结果的有效性,以发现更加符合实际需求的典型工艺路线。然而,其所提 AP 算法的聚类结果受到参考度(p)与阻尼系数(λ)的直接影响,且人为设置这两个参数较为困难。因此,如何寻找合适的参考度和阻尼系数是运用 AP 算法求解相关问题的一个关键点。

1.3 本文主要研究内容

为了有效解决上述问题,本文提出一种多维工艺信息融合的典型网状工艺路线自动化发现方法,其中的主要工作和创新点如下:

(1)本文详细分析了计算网状工艺路线相似度时需要考虑的工艺信息需求,并基于此构建了面向网状工艺路线的不同相似度情况案例库。

(2)基于上述分析,本文设计了一种面向网状工艺路线的新型综合相似度度量方法。该方法创新性地将网状工艺路线中的先后顺序关系和工序相似度信息量化问题转化为工序对的加权二分图最大匹配问题并通过库恩-曼克尔斯(Kuhn Munkres, KM)算法^[36]计算了最大匹配权重。此外,所提方法还以 Jaccard 相似度^[37]量化了共同工序数、以可选工艺路线数目量化了结构复杂度,并通过主成分分析(Principal Component Analysis, PCA)整合成最终的综合相似度。

(3)由于考虑了聚类有效性的典型工艺路线发现问题的求解难度陡增,为了解决 AP 算法输入参数难以人为调节的问题,本文将火鹰优化算法(Fire Hawk Optimizer, FHO)^[38]引入到了 AP 算法中,并提出了 FHO-OAP 和 FHO-IAP 两种算法变体,以实现 p 和 λ 的自动调整,避免人为设置参数的困难和主观性,从而在聚类效果和聚类有效性中取得平衡,获得更符合实际需求的典型网状工艺路线。

基于此,本论文的框架结构如下:

第 1 章为绪论,依次介绍了多维工艺信息融合的典型网状工艺路线自动发现方法的研究背景及意义,并且对国内外相关研究现状进行了详细的分析,从而引出本文的研究方法与研究内容。

第 2 章详细介绍相关理论知识。对于网状工艺路线的综合相似度设计,系统性地分析了网状工艺路线的综合相似度设计内在的工艺信息需求,并基于分析结果构建了包含不同相似情况的案例库。对于考虑聚类有效性的典型工艺路线发现过程,本章主要研究了包含数量软约束和半径软约束的网状工艺路线聚类问题,考虑到传统 AP 算法在求解这类复杂问题时难以人为合理设置算法参数,本章提出引入 FHO 算法来迭代优化 AP 算法的参考度(p)与阻尼系数(λ)。

第3章主要研究多维工艺信息融合的新型网状工艺路线相似度设计。根据第2章所提的四种不同类型的工艺信息，以双结构矩阵和KM算法量化了先后顺序关系和工序间相似度、以Jaccard相似度量化了共同工序数、以可选路线数目量化了结构复杂度。同时，考虑三个中间相似度之间的相关性采用PCA方法将这三者合成为一个综合相似度。最后通过仿真实验验证了该方法的优越性。

第4章主要研究考虑聚类有效性的典型网状工艺路线发现方法。本章提出了FHO-OAP和FHO-IAP两种变体算法，以期在求解问题时能在聚类结果和两个软约束之间取得平衡，发现更加符合实际需求的典型网状工艺路线。最后通过仿真实验与对比体现出所提方法的有效性。

第5章对本文的研究内容进行了概括性的总结，并且指出当前研究工作的不足之处，为未来研究方向指出了明确的道路。

为了清晰展示本文的研究思路与总体脉络，以图1-1来简述本文的总体章节框架。

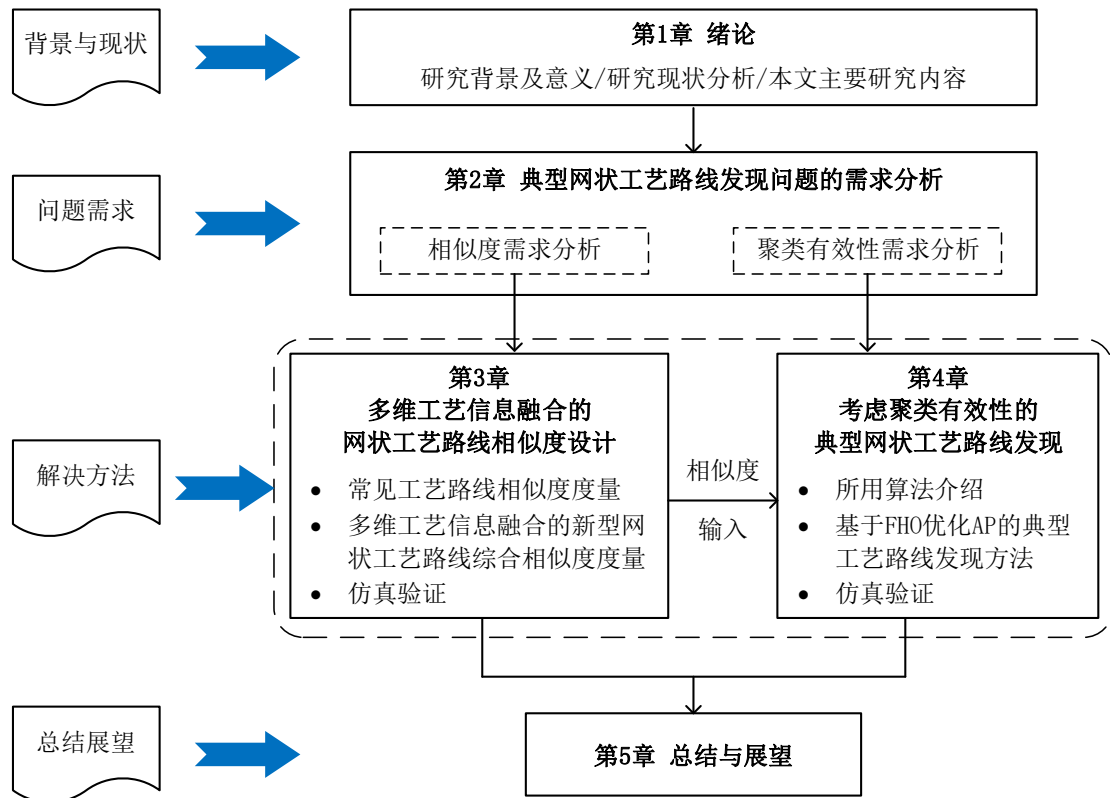


图 1-1 论文结构图

第2章 典型网状工艺路线发现问题的需求分析

2.1 引言

传统的工艺路线规划方式在面对日益复杂多样的产品需求时，逐渐暴露出效率低下、灵活性不足等问题。在此形势下，典型工艺路线的发现成为工艺规划领域的研究热点。通过挖掘具有代表性的典型工艺路线，不仅可以为新产品的工艺规划提供参考模板，还能促进工艺知识的积累与传承，推动工艺规划过程朝着标准化、智能化方向发展。

在典型工艺路线发现的研究中，网状工艺路线相似度分析以及典型网状工艺路线的发现是两个关键且紧密相关的环节。准确度量网状工艺路线的相似度，有助于在大量的工艺路线数据中找到相似案例，为后续的典型工艺路线提取提供基础；而在发现典型网状工艺路线过程中，聚类有效性问题至关重要，它直接关系到所获取的典型工艺路线能否真实反映生产实际并具有广泛的应用价值。

鉴于此，本章将围绕典型工艺路线发现展开深入探讨。首先阐述其总体框架与思路，接着详细剖析网状工艺路线相似度面临的问题难点，包括信息需求分析以及案例库构建。然后针对典型网状工艺路线发现的问题难点，从聚类有效性需求分析以及相应的数学描述方面进行研究。最后通过本章小结对研究内容进行总结，旨在为典型工艺路线发现的研究提供全面且系统的理论支持与实践指导。

2.2 典型工艺路线发现的总体框架与思路

在制造业智能化转型中，典型工艺路线的发现与提取是优化生产流程、提升制造效率的关键任务。典型工艺路线的发现主要包含两个核心步骤：相似度设计和聚类分析。这两个步骤紧密关联，共同构成了典型工艺路线发现的核心框架。首先，相似度设计旨在建立工艺路线之间的相似性评估体系。通过设计网状工艺路线的相似度计算方法，能够量化不同工艺路线之间的相似程度。这一过程依赖于丰富的案例库和工艺数据，为后续的聚类分析提供可靠的数据基础。相似度设计的核心目标是准确捕捉工艺路线的特征，确保评估结果的科学性和实用性。

其次,聚类分析是基于相似度设计的结果,对网状工艺路线进行分组和归类。通过采用先进的聚类算法,将具有相似特征的工艺路线归并为同一类别,从而识别出具有广泛适用性和高效特征的典型工艺路线。聚类分析不仅能够发现潜在的典型工艺路线,还能为工艺优化提供数据支持。这两个步骤之间存在紧密的关联性:相似度设计是聚类分析的前提,它为聚类算法提供了量化的相似性度量标准;而聚类分析则是相似度设计的应用,通过将相似度计算结果转化为具体的工艺路线分组,最终实现典型工艺路线的发现。两者相辅相成,共同解决了典型工艺路线发现问题。

2.3 网状工艺路线相似度的问题难点

2.3.1 网状工艺路线相似度设计的信息需求分析

在设计网状工艺路线相似度之前,需要分析网状工艺路线中包含了怎样的信息。换言之,是要分析出什么样的信息可以体现两条工艺路线的细微差别。通过已有相关研究的分析^[11,28],本文结合网状工艺路线的实际需求,认为网状工艺路线中包含着四种信息:先后顺序关系、工序间相似度、共同工序数和结构复杂度,如图 2-1 所示。这些信息尽可能地决定了一条工艺路线的唯一性,也能通过这些信息来度量不同工艺路线的相似性或差异性,从而为制定更有效的生产计划和优化生产流程提供重要参考。因此,深入了解和分析量化这些多维工艺信息可以更合理且有效地设计典型工艺路线发现过程中的网状工艺路线相似度度量。

(1) 先后顺序关系分析

工艺路线中工序的先后顺序关系是生产精度和工艺逻辑顺序的基础,它反映了工序间的依赖和顺序关系,确保零件生产过程中工艺上的合理性。工艺规划的本质是综合考虑时间、资源、技术、工艺限制等多重要素,来指定各个工序的优先级,以提高效率、降低成本。相关研究已表明^[39],同类型产品通常具有相似的工序先后顺序关系,因为它们的工艺需求大体相同。

在研究线性工艺路线时,先后顺序关系通常通过最长公共子串或最长公共子序列来衡量^[40]。然而,这两种方法仅适用于线性工艺路线,因为网状工艺路线中存在多个可选路线,可以看成是多条线性工艺路线的集合。因此,在网状工艺路

线中使用最长公共子串或最长公共子序列进行相似度比较将变得复杂且不准确，可能会造成部分公共子串或子序列的重复计算，从而影响相似度度量的准确性。此外，还应考虑工序间的相似性，否则先后顺序关系将仅限于共同工序，忽略相似工序的先后顺序信息。

本文采用网状工艺路线的双结构矩阵和 KM 算法来寻找最大权重匹配^[36]，从而可以同时衡量工序相似度和工序的先后顺序信息。具体来说，双结构矩阵是通过将工序相似度和工序的先后顺序关系转化为二部图的方式来表示。在这个二部图中，一部分节点代表工序，另一部分节点代表工序的先后顺序关系。通过计算节点之间的相似度，可以得到一个权重矩阵，表示工序之间的相似程度。之后，利用 KM 算法来寻找最大权重匹配，从而确定最佳的工序优先关系。通过这种方法，本文不仅可以考虑工序之间的相似度，还可以考虑工序之间的先后顺序关系，从而更全面地衡量网状工艺路线之间的关联性。

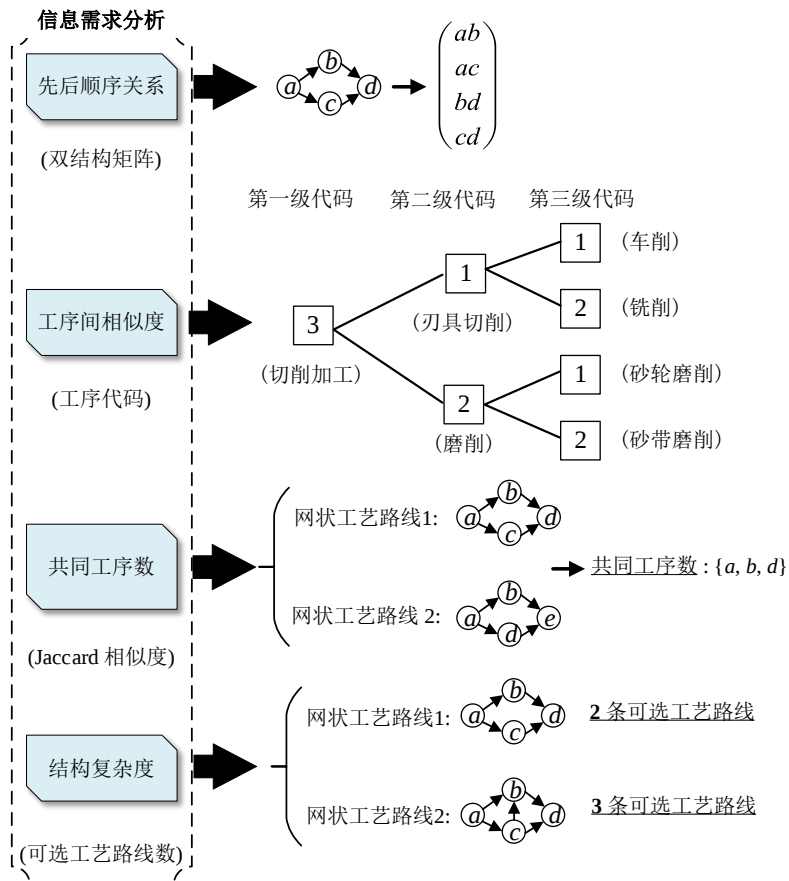


图 2-1 网状工艺路线的信息需求分析

（2）工序间相似度

工序作为工艺路线的最小单位，其本身包含了大量的工艺信息，如加工方法、精度要求、处理方法等。在以往的研究中，只考虑了工序完全相同和完全不同两种情况，忽略了工序中嵌入的信息，即没有考虑工序间本身的相似度。

在本文中，本文遵循 Liu 和 Zhang 等人^[26]等人的研究，基于国家标准使用三级代码来描述每个工序中与流程相关的信息。如图 2-1 所示，第一级代码按工序专业划分，第二级代码按第一级代码的工序特点划分，第三级代码的划分按机械行业标准 JB/T5992.2~5992.10 的规定划分。然后，根据这些代码即可更为合理地计算不同工序间的相似度。

（3）共同工序数

在上述基于双结构矩阵和 KM 算法度量工序相似度和先后顺序关系的过程中，实际已经包含了共同工序的信息，因为共同工序本质上就是工序间相似度为 1 的工序对。但上述度量方法不仅考虑了共同工序信息，也同时考虑了这些工序的先后顺序信息。换言之，如果相同工序的顺序不同，那么上述方法则认为他们的在先后顺序关系上的相似度为 0。

为了获取共同工序的独立信息，就需要仅计算共同工序的数量，将这部分信息与工序的先后顺序信息解绑。为了实现这一目标，本文采用 Jaccard 相似度^[37]来衡量工序之间的关联性。Jaccard 相似度是一种衡量两个集合之间相似程度的方法，它考虑了两个集合的交集和并集的比值。在这里，本文将工序看作集合，共同工序的数量即为交集的大小，工序的总数量即为并集的大小。

（4）结构复杂度

网状工艺路线的结构复杂度是评价其复杂程度和结构信息的重要指标之一，它反映了网状工艺路线中可选工艺路线的数量，即工艺路线的多样性和柔性。通俗地说，如果一条网状工艺路线有更多的可选路线那么它的结构复杂度就更高。这意味着在生产过程中对应零件有更多的工艺选择和灵活性，同时也意味着对于可能的变化或错误有更好的容错性。

在本文中，采用可选工艺路线的数量来衡量网状工艺路线的结构复杂度。如此，线性工艺路线的结构复杂度为 1，可以看成是最简单的一种网状工艺路线。

2.3.2 网状工艺路线相似度案例库构建

在分析完网状工艺路线相似度设计的信息需求后，本节综合上述分析结果，总结出网状工艺路线的所有不同相似情况。如表 2-1 所示，本文从工序、结构、顺序三个方面入手，考虑每个方面的可能情况，构建了一个面向网状工艺路线的相似情况案例库。其中，图 2-1 所示的共同工序信息被认为是工序相似度为 1 的特殊情况。这些不同的相似情况表明两条网状工艺路线间包含着不同的公共信息，从而导致不同的相似程度。

表 2-1 三种不同相似情况

三个方面	不同情况
工序	a. 路线间所有工序相同； b. 路线间所有工序相似； c. 路线间部分工序相同； d. 路线间部分工序相似； e. 路线间所有工序不同。
结构	a. 可选路线数量相同； b. 可选路线数量较多； c. 可选路线数量较少。
顺序	a. 工序顺序相同； b. 工序顺序不同； c. 工序顺序部分相同。

根据表 2-1，以原始工艺路线（记为 OG）为基准，对其进行不同情况下的适当修改，可以建立 39 种不同的相似度情况，如图 2-2 所示。一目了然，案例 1 的相似性肯定是最高的，因为所有的信息都是相同的。而案例 38 和 39 是最差的，因为这两个案例只有一点结构相似性。此外，由于案例 37、38、39 与原始工艺路线 OG 没有任何共同或相似的工序，因此讨论这三个案例的先后顺序信息是没有意义。换言之，这三个案例只能从工艺路线结构信息上谈论一定的相似性。

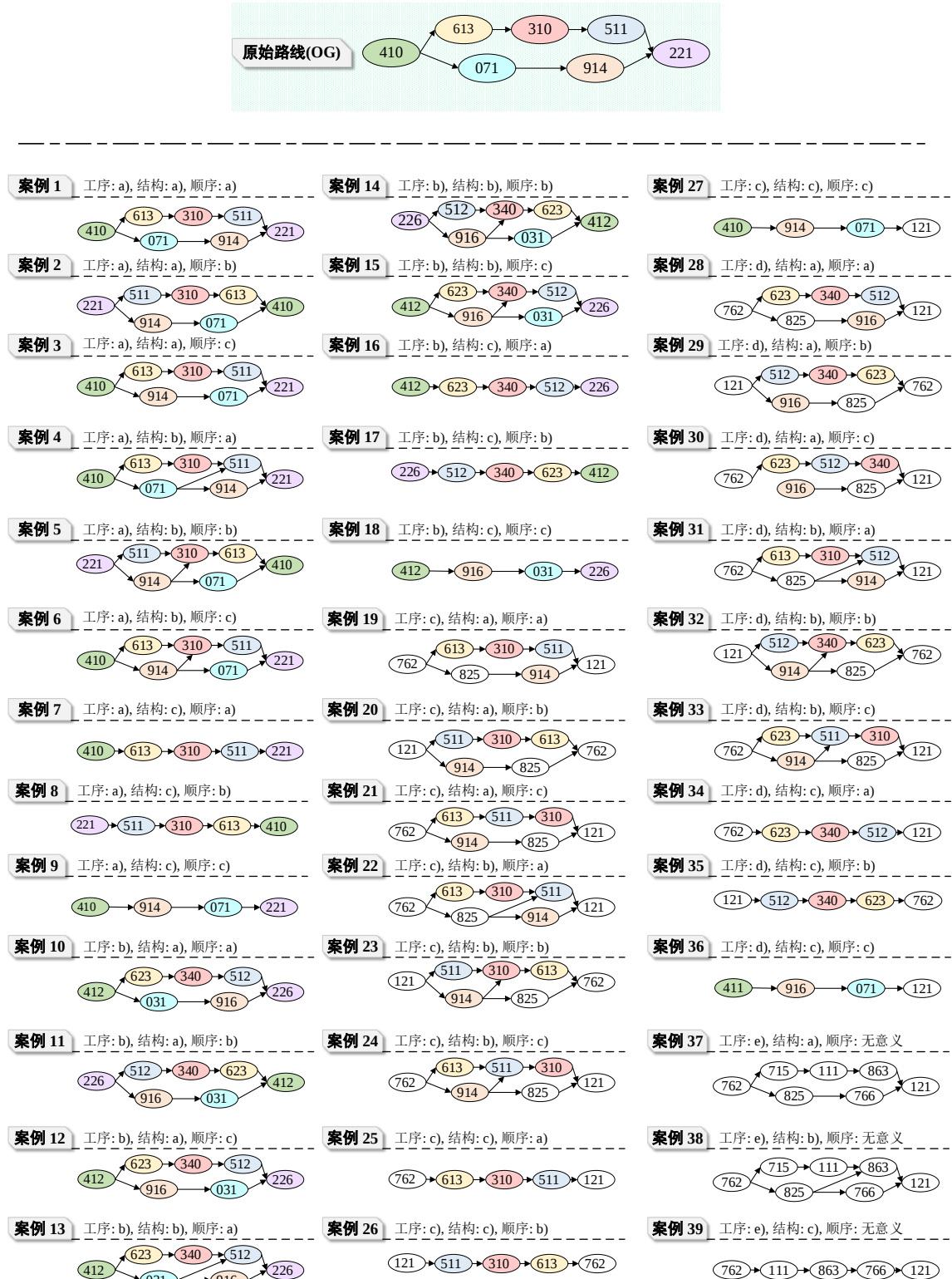


图 2-2 网状工艺路线的不同相似情况

2.4 典型网状工艺路线发现的问题难点

2.4.1 典型网状工艺路线的聚类有效性需求分析

典型工艺路线聚类与一般聚类问题在数据特征、目标和聚类结果等方面存在显著区别。

首先，从数据特征来看，工艺路线数据通常具有多维、时序性和依赖性的特点。这种多维性体现在工艺数据的多个维度上，包括设备参数、操作条件、产能指标等，每个维度可能包含复杂且动态的变化信息。时序性则意味着工艺步骤具有严格的顺序，某些操作必须依赖于先前步骤的完成结果，而这种时间依赖性直接影响了整个工艺流程的运行效率。工艺数据的依赖性还表现在设备之间的关联性和步骤间的逻辑关系上，例如一个设备的输出可能直接成为下一个步骤的输入，这种特性使得工艺路线数据具备较强的整体性和连贯性。相比之下，一般聚类问题的数据往往是静态的，缺乏时间序列信息，样本之间通常是独立的，没有明显的因果关系。这种独立同分布的假设使得普通聚类方法在处理工艺路线数据时面临较大的局限性。

其次，在聚类目标上，工艺路线聚类具有更为复杂和明确的业务导向。其目标不仅仅是发现数据的分布模式，还需要深度揭示不同工艺路径之间的相似性和差异性。这种相似性和差异性往往包含生产效率、工艺成本和质量稳定性等方面的业务意义。通过工艺路线聚类，能够帮助制造企业识别出效率较高的典型工艺路径，找出存在问题或优化空间的异常路径，从而为工艺优化和流程改进提供决策支持。而普通聚类问题的目标通常集中于将样本划分为若干组，强调的是样本分布结构的整体性和数学上的聚类效果，较少关注结果的业务含义和实际应用。

最后，在聚类结果上，是典型工艺路线聚类与一般聚类问题最本质的区别。典型工艺路线聚类的结果是从每个聚类簇中选取一个真实存在的工艺路线作为该簇的代表，即典型工艺路线。这个典型工艺路线不仅需要能够有效代表该聚类簇中大部分的工艺信息，还需确保不同聚类簇之间的差异性尽可能大。因此，在典型网状工艺路线的聚类过程中，往往需要引入额外的约束条件，以保证聚类结果的有效性。

如果典型工艺路线聚类未能充分考虑聚类的有效性,可能会对实际应用产生多方面的负面影响。

首先,聚类结果可能无法准确代表聚类簇的特征。典型工艺路线聚类的核心目标是从每个聚类簇中选取一个真实存在的工艺路线作为代表,用以反映该簇大部分工艺路线数据的关键特征。如果未能保证聚类的有效性,选出的典型工艺路线可能缺乏足够的代表性,从而无法准确描述簇内工艺路线数据的整体特征。这将导致企业无法通过这些代表性工艺路线深入理解现有工艺模式。

其次,不同聚类簇之间的区分度可能不足。有效的工艺聚类要求不同簇之间具有显著的差异性,以便清晰地分离不同类型的工艺路径。如果聚类结果未能体现这一点,不同簇之间的边界可能模糊,导致簇内和簇间的工艺路线高度相似甚至重叠。这种情况会削弱聚类结果的指导意义,使企业难以基于这些结果制定针对性的工艺优化策略。

最后,聚类结果可能失去业务上的指导价值。如果未能关注有效性,聚类过程可能仅停留在数学分组的层面,而未能反映工艺路径的业务需求和实际意义。例如,选出的典型工艺路线可能无法准确揭示高效工艺流程,也难以帮助识别异常路径或优化空间,从而影响工艺优化、成本控制和质量提升的实际应用。

2.4.2 考虑聚类有效性的典型网状工艺路线发现问题数学描述

典型工艺路线发现问题的核心在于通过聚类找到中心点,虽然线性工艺路线的相似度度量不能直接用于网状工艺路线中,但关于典型工艺路线的发现问题的,无论工艺路线是线性还是网状,都是根据工艺路线相似度的差异性进行适当聚类,故聚类方法对于网状和线性的工艺路线是通用的。考虑到典型工艺路线聚类是一个实际问题,所以需要考虑聚类有效性,具体而言,包括数量约束和半径约束^[35]。

数量约束是指聚类结果中每个聚类簇中的工艺路线数目应当超过最小的工艺路线数目 n ,而半径约束是指聚类结果中每个聚类簇的半径应当小于最大的聚类簇半径 ε 。前者用于保证聚类簇中包含有足够的工艺数据,否则这一聚类簇中生成的典型工艺路径就不具有较高的普适性。而后者则是用来保证发现的典型工艺路线不会因为聚类簇中的工艺信息过多而难以很好的代表簇中的其他工艺路

线。这两个约束条件提出了聚类有效性的概念，从而使得发现的典型工艺路线更符合实际需求。

Wang 和 Zhang 等人^[41]最先在典型工艺路线发现问题中引入了数量约束和半径约束来提高聚类结果的有效性和优化聚类数目。但他们将这两个约束作为硬约束处理，这可能导致聚类结果不够灵活，无法很好地适应不同数据集的分布变化。相比之下，文献[35]通过传统 AP 算法(Original AP, OAP)中加入了半径软约束和数量软约束，设计出了改进的 AP 算法(Improved AP, IAP)将约束参数的敏感度降低，从而提升聚类结果的有效性。然而，徐的工作也存在问题。引入了半径软约束和数量软约束的典型工艺路线发现问题相较于传统问题更加复杂，需要在聚类结果和软约束罚值间寻求平衡，而单靠人为设置合理的参考度(p)与阻尼系数(λ)显然是非常困难的，极易使 IAP 算法陷入局部最优，难以保证所得结果的全局最优性。

为了改进这一问题，本文将 FHO 算法与 AP 算法相结合，提出基于 FHO 优化 AP 的聚类算法，用来自动选择最优的 p 和 λ ，以保证 AP 算法聚类结果的最优性，从而在考虑聚类有效性的同时生成最具代表性的典型工艺路线。

考虑聚类有效性的典型网状工艺路线发现问题的数学描述如下：对于给定的一个含有 N 条网状工艺路线的集合 $\mathbf{X} = \{x_1, \dots, x_N\}$ ，定义 $\mathcal{S} = \{sim(x_i, x_j)\}$ 为 $\mathcal{X}$ 中所有网状工艺路线之间的相似度集。其中 $sim(x_i, x_j)$ 是根据 3.3 节提出的多维工艺信息融合的新型网状工艺路线综合相似度计算得到。定义 $q_e \geq 0$ 和 $q_r \geq 0$ 分别是数量软约束和半径软约束的罚值。聚类算法的目标是找到中心集合 $O = \{o_1, \dots, o_K\}$ 使得如下目标函数 $netsim$ 最大化：

$$netsim = \sum_{i=\{1,2,\dots,N\}} s(x_i, O_{xi}) - q_e |X_e| - \sum_{x_i \in X_r} \left(\frac{\varepsilon - s(x_i, O_{xi})}{\varepsilon} q_r \right), \quad (2-1)$$

其中， O_{xi} 为第 i 条工艺路线 x_i 的典型工艺路线（即中心）， $\mathcal{X}_e$ 和 $\mathcal{X}_r$ 分别是违反数量约束和半径约束的工艺路线集合， ε 是网状工艺路线有效聚类群的最大半径。 $netsim$ 表示各工艺路线与其对应中心的相似度和及各软约束的罚值的累加。

2.5 本章小结

本章研究了典型工艺路线的发现。主要内容包括总体框架的构建、网状工艺路线相似度的分析与案例库的建立,以及典型网状工艺路线发现问题的数学描述。首先,构建了总体框架,明确了典型工艺路线发现的目标、思路和步骤。框架包括需求分析、网状工艺路线相似度评估、聚类算法的应用和结果验证,提供了完整的理论支持和实践指导。然后,针对网状工艺路线相似度的设计,分析了先后顺序关系、工序间相似度、共同工序数和结构复杂度等信息。采用了基于双结构矩阵和 KM 算法的相似度度量方法,并构建了案例库,为后续研究提供了基础。最后,研究了典型网状工艺路线聚类问题。根据数量约束和半径约束,提出了改进的 AP 算法和 FHO 优化策略,解决了聚类有效性的问题,并通过数学描述明确了模型。本章为后续典型工艺路线发现提供了理论和方法。

第3章 多维工艺信息融合的网状工艺路线相似度设计

3.1 引言

作为生产制造过程中的核心要素之一，工艺路线的合理性直接影响到生产效率、资源利用率和产品质量。在传统制造环境中，线性工艺路线的相似度计算方法已经相对成熟，但随着网状工艺路线在复杂制造场景中的应用日益广泛，如何有效地衡量和比较网状工艺路线的相似性，成为研究的关键难题。

本章系统探讨了工艺路线相似度的相关理论与方法。首先，分析了常见的工艺路线相似度度量方法及其适用范围，包括基于序列比较和距离公式的经典算法。在此基础上，针对网状工艺路线的复杂结构与特性，提出了一种结合多维工艺信息的新型综合相似度度量方法，通过工艺间相似度、先后顺序关系、结构复杂度等维度进行综合分析和计算，来保证所得相似度度量的合理有效。此外，还对提出的方法与传统方法进行了性能比较与验证。

3.2 常见工艺路线相似度度量

3.2.1 基于最长公共子序列算法的相似度

Askin 和 Zhou^[42]提出了一种基于工序之间 LCS 的相似系数，用于线性工艺路线。他们构建了一个树，其中包含两个部分的工序序列中常见的所有可能的工序子序列。此外，非支配匹配的概念已被应用于减小树的大小。两个运算序列 x 和 y 之间的相似系数定义 S_{xy} 为：

$$S_{xy} = \max \left\{ \frac{|LCS_{xy}|}{|x|}, \frac{|LCS_{xy}|}{|y|} \right\}, \quad (3-1)$$

其中， LCS_{ij} 是 x 和 y 之间的最长公共子序列， $|x|$ 是序列 x 中的工序数。

3.2.2 基于合规指数的相似度

Ho 和 Lee 等人^[43]提出了一个工序序列相似系数，该系数基于产品序列中正向和反向流动序列依次或旁路的工序数量。工序序列相似系数定义为两个兼容指标之和除以乘积中操作次数的两倍。产品 x 的操作序列相似系数可以表示为：

$$CO_{xy} = \frac{(CF_x + CB_x)}{2 \times N_x} \quad (3-2)$$

其中， CO_{xy} 是与部分 y 合并的部分 x 的工序序列相似系数， CF_x 是正向兼容索引， CB_x 是反向兼容索引， N_x 是部分 x 操作序列中的工序次数。在寻找相似系数的过程中，假设部分 x 在其工序序列中具有两个部分中最少的工序数； CO_x 的值在 0 和 1 之间。

3.2.3 多维工艺路线最长公共子序列检索算法

罗年猛和李雄^[44]等人提出了适合于检索过程中工艺路线相似度的多维工艺路线最长公共子序列算法。

该算法首先将工艺路线表示为， $X_m = (x_1, x_2, \dots, x_i, \dots, x_m)$ ，其中 x_i 为工序，将工序编码表达为大类、中类、小类、细分类的层次结构。工序 x_i 编码为 $x_i = L_{i1}L_{i2}L_{i3}(C_{i1}, Q_{i1}, C_{i2}, Q_{i2}, \dots, C_{ik}, Q_{ik}, \dots, C_{it}, Q_{it})$ 。其中， L_{i1}, L_{i2}, L_{i3} 分别表示当前工序 x_i 的大类、中类、小类的编码， C_{ik} 表示工序中第 k 类参数的编码； Q_{ik} 表示第 k 类参数对工序的影响权值。

对于两条工艺路线 $X_m = (x_1, x_2, \dots, x_i, \dots, x_m)$ 和 $Y_n = (y_1, y_2, \dots, y_j, \dots, y_n)$ ，其中工序 x_i 和 y_j 的编码形式依次为： $x_i = L_{i1}L_{i2}L_{i3}(C_{i1}, Q_{i1}, C_{i2}, Q_{i2}, \dots, C_{ik}, Q_{ik}, \dots, C_{it}, Q_{it})$ ， $y_j = L_{j1}L_{j2}L_{j3}(C_{j1}, Q_{j1}, C_{j2}, Q_{j2}, \dots, C_{jk}, Q_{jk}, \dots, C_{jt}, Q_{jt})$ 。当两工序的前三位编码一致时，有 $Q_{ik} = Q_{jk}$ ，或 Q_{ik} 、 Q_{jk} 中有一个为 0。

对于上述两条工艺路线，罗年猛和李雄^[44]采用多维工艺路线最长公共子序列算法计算了它们之间的相似度 SIM ：

$$SIM(X_m, Y_n) = \frac{MLC'(X_m, Y_n)}{\max(m, n)}, \quad (3-3)$$

其中, $MLC'(X_m, Y_n)$ 是两条工艺路线之间的多维最长公共子序列的长度, 如公式 (3-4) 所示:

$$MLC'(X_m, Y_n) = \begin{cases} 0 & i=0 \text{ 或 } j=0 \\ MLC'(X_{i-1}, Y_{j-1}) + 1 & sim(x_i, y_j) \geq \theta, \\ \max[MLC'(X_{i-1}, Y_j), MLC'(X_i, Y_{j-1})] & \text{其他} \end{cases} \quad (3-4)$$

其中, θ 是预先设定的工序相似度阈值; $sim(x_i, y_j)$ 是当前工序 x_i 和 y_j 的相似度:

$$sim(x_i, y_j) = \begin{cases} -\infty & L_{i1} \neq L_{j1} \text{ 或 } L_{i2} \neq L_{j2} \text{ 或 } L_{i3} \neq L_{j3} \\ \theta & Q = 0 \\ \alpha & \text{其他} \end{cases}, \quad (3-5)$$

其中, Q 为工序细化编码的影响权值之和, $Q=0$ 表示至少有一道工序的细化编码没有选取; α 是一般相似度。

$$Q = \sum_{k=1}^t \min\{Q_{ik}, Q_{jk}\}, \quad (3-6)$$

$$\alpha = \frac{1}{Q} \sum_{k=1}^t \delta(C_{ik}, C_{jk}) \times \min(Q_{ik}, Q_{jk}), \quad (3-7)$$

$$\delta(C_{ik}, C_{jk}) = \begin{cases} 1 & C_{ik} = C_{jk} \\ 0 & \text{其他} \end{cases}, \quad (3-8)$$

多维工艺路线最长公共子序列检索算法主要用于分析和优化工艺路线中的工序顺序。该算法假设工艺路线之间的关系是线性的, 即工序之间存在明确的前后顺序。然而, 网状工艺路线具有不同的特点, 其工序间可能存在多条并行路径、交叉依赖以及可能的循环结构。这些复杂的关系使得网状工艺路线难以通过简单的线性序列方式来确定最长公共子序列。

网状工艺路线的工序依赖关系并非单向或固定, 其特点在于工序间可能存在多个并行或交叉的关系。由于这些非线性特征, 多维工艺路线最长公共子序列检索算法无法直接应用于网状工艺路线。因此, 需要考虑专门适配网状工艺路线的相似度度量方法来处理工序间的复杂依赖和多路径特性。

3.2.4 采用距离公式计算工艺路线相似度

Liu 和 Zhang 等人^[26]运用曼哈坦(Manhattan)距离^[45]计算工序间相似度,运用欧氏距离^[45]计算工艺路线间的相似度,并以相异度矩阵表示工艺路线的相异度。相异度指的是两条工艺路线之间不相似的程度。

用具有 n 行 p 列的数据矩阵表示工艺路线集合,如式(3-9)所示:

$$X_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p}, \quad (3-9)$$

其中,该矩阵中每一行代表一条工艺路线, x_{ij} 表示第 i 条工艺路线的第 j 道工序的工序编码。 n 表示工艺路线总数, p 表示工艺路线中工序数目的最大值。

(1) 曼哈坦距离计算工序间相似度

曼哈坦距离是一种适用于计算编码距离的距离度量方法,针对工序编码距离计算的特点,在曼哈坦距离公式中加入权值并将距离值归一化的公式(3-10):

$$d_1(x_{ik}, x_{jk}) = \|x_{ik} - x_{jk}\|_1 = \frac{1}{b} \cdot \sum_{m=1}^l C_m |x_{ikm} \wedge x_{jkm}|, \quad (3-10)$$

其中, $b = \sum_{m=1}^l C_m$, $d_1(x_{ik}, x_{jk})$ 表示工艺路线 i , j 中第 k 道工序的距离, l 表示参与计算编码的长度; x_{ikm} 表示工艺路线 i 中第 k 道工序的前 m 位编码, $m=1, 2, 3, 4$ 分别代表工序编码的大类、中类、小类和细分类。 $x_{ikm} \wedge x_{jkm}$ 表示 x_{ikm} 和 x_{jkm} 的异或运算; C_m 表示加工方法各个类对工序相似度的影响系数。

(2) 欧氏距离计算工艺路线间的相异度

用 $d_2(x_i, x_j) \in [0, 1]$ 表示工艺路线 x_i 和 x_j 之间的相异度,式(3-11)是归一化的欧氏距离计算公式:

$$d_2(x_i, x_j) = \|x_i - x_j\|_2 = \left[\frac{1}{p} \sum_{k=1}^p \|x_{ik} - x_{jk}\|_1^2 \right]^{\frac{1}{2}}, \quad (3-11)$$

其中, $\|x_{ik} - x_{jk}\|_1$ 表示工序 x_{ik} 和工序 x_{jk} 的距离,也就是工序间相似度, p 代表的是工艺路线中工序数目的最大值。

(3) 相异度矩阵

通过上述对工序和工艺路线间距离的计算,可以计算出对应于式(3-9)的数据矩阵的相异度矩阵,如式(3-12)所示:

$$D_{n \times n} = \begin{bmatrix} 0 & & & \\ d_2(x_2, x_1) & 0 & & \\ \vdots & \vdots & 0 & \\ d_2(x_n, x_1) & d_2(x_n, x_2) & \cdots & 0 \end{bmatrix}_{n \times n}, d_2(x_i, x_j) \in [0, 1] \quad (3-12)$$

曼哈顿距离和欧氏距离在计算工序和工艺路线相似度时,通常依赖于线性结构的特性。曼哈顿距离用于计算工序间的相似度,其基础假设是工序可以通过固定的坐标差值来衡量。而欧氏距离用于计算工艺路线间的相似度,通常通过点与点之间的直线距离来表示。这种方法隐含的前提是工艺路线具有线性、单向的结构,即工序间的依赖关系是明确且固定的。

然而,网状工艺路线具有复杂的非线性结构,包括多条并行路径、交叉依赖以及可能的循环。这种复杂性使得网状工艺路线的工序关系无法简单地用点对点的距离来准确表示。此外,网状结构中的工序关系并非单一线性路径,而是多个交织关系的综合,导致传统基于线性距离的计算方法无法有效处理这种多维依赖。因此,曼哈顿距离和欧氏距离在网状工艺路线的相似度计算中失去适用性,仅能用于线性工艺路线中。

3.3 多维工艺信息融合的新型网状工艺路线综合相似度量

如 2.3.1 节所示,一个合理的网状工艺路线度量应该综合考虑所分析提及的四种信息。只有通过对这四种信息的有效量化与综合,才能更好地捕捉不同网状工艺路线之间的细微差异,从而得到灵敏度更高且细粒度的综合相似度,也才能保证最后所得典型工艺路线的质量。为此,本章提出了一种面向网状工艺路线的新型综合相似度量,通过对多维工艺信息的合理量化与综合,来保证相似度量度的有效性,其总体设计框架如图 3-1 所示。

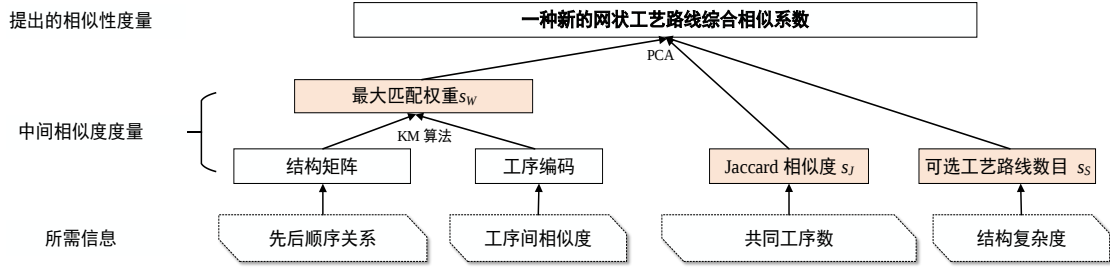


图 3-1 所需信息和相似度量之间的层次关系

3.3.1 工序间相似度量

结合 Liu 和 Zhang 等人^[26]的研究，基于国家标准将工序表示为三位数的编码形式。这种表示方式本质上是将所有的工序归类到不同层次的类别中，每一层都专注于嵌入在工序本身中的与工艺过程相关的信息。如图 2-1 所示，用 311 代表车削，用 322 表示砂带磨削等。接着，根据工序三级编码中不同层级的工艺信息，通过同或运算(XNOR)计算工序间相似度：

$$s_o(o_x, o_y) = (\sum_{z=1}^3 o_x^z \odot o_y^z) / 3 \quad (3-13)$$

其中， o_x 和 o_y 分别是两个工序， o_x^z 和 o_y^z 分别是这两个工序的前 z 位编码。根据该定义，举例计算： $s_o(410, 412) = \frac{(1+1+0)}{3} = \frac{2}{3}$ ， $s_o(613, 623) = \frac{(1+0+0)}{3} = \frac{1}{3}$ 。

3.3.2 先后顺序关系度量

先后顺序关系反映的是不同工序在工艺上的前后依赖关系。但常规度量该信息的前提是两个工序是完全相同的，否则讨论先后顺序关系就毫无意义。然而，由于工序之间具有不同的相似度(见式(3-13))，因此度量先后顺序关系显然不能忽略工序间的相似性。如图 3-1 所示，本文采用双结构矩阵和 KM 算法相结合的方式，将这两种工艺信息综合考虑和度量。

如图 3-2 所示，本文将网状工艺路线中各工序的先后顺序关系表示为双结构矩阵。具体而言，是从网状工艺路线中提取出所有相邻的工序对，并将其排列成两列矩阵形式。这样，可以有效地将网状工艺路线从网络图的形式转换成专门描述其中先后顺序的计算机可识别形式。同时，也保留了网状工艺路线的先后顺序

信息。在此基础上,定义两个相邻工序对(分别记为 $opr_i = (o_{i1}, o_{i2})$ 和 $opr_j = (o_{j1}, o_{j2})$)的相似度为:

$$s_{pr}(opr_i, opr_j) = \sum_{k=1,2} 0.5 \times s_o(o_{ik}, o_{jk}), \quad (3-14)$$

其中, $s_o = (o_{ik}, o_{jk})$ 表示的工序 o_{ik} 与 o_{jk} 工序之间的相似度(见式(3-13))。而 k 为 1 时, 计算的是工序对首部之间的相似度; k 为 2 时, 计算的是工序对尾部之间的相似度。

得到网状工艺路线的双结构矩阵之后, 本文将工艺路线中先后顺序关系和工序间相似度的量化问题转化为加权二分图的最大匹配问题。其中, 网状工艺路线双结构矩阵中的各相邻工序对就是相应的匹配对象, 而匹配权重则由式(3-14)计算得到。

KM 算法由 Harold Kuhn 和 James Munkres^[46]首次提出, 该算法的核心思想是通过不断调整指派方案来逐步优化总成本, 直至找到最优解。该算法基于匈牙利方法的思想, 利用匈牙利图来描述指派问题, 并采用增广路径的方式来实现最优解的搜索。KM 算法相较于匈牙利算法的不同在于它引入了权值这一概念因此非常适合用于解决加权二分图的最大匹配问题。

基于上述原因, 本文选择 KM 算法来求解两个网状工艺路线双结构矩阵的最大权重匹配问题, 并将得到的最大匹配权重作为同时考虑先后顺序关系和工序间相似度的相似度度量 s_w :

$$s_w(nos_m, nos_n) = \frac{\sum_{\max\{|dsm_m|, |dsm_n|\}} s_{pr}(opr_i, opr_j)}{\max\{|dsm_m|, |dsm_n|\}} \quad (3-15)$$

在(3-15)式中, nos_m 和 nos_n 为两条网状工艺路线, 而 $|dsm_m|$ 和 $|dsm_n|$ 是 nos_m 和 nos_n 对应双结构矩阵中相邻工序对的个数。

例如, OG 与案例 11 的最大匹配权值如图 3-2 所示, 其最大匹配权重为 $s_w(OG, \text{案例11}) = 8/21$ 。

OG的结构矩阵 案例11的结构矩阵

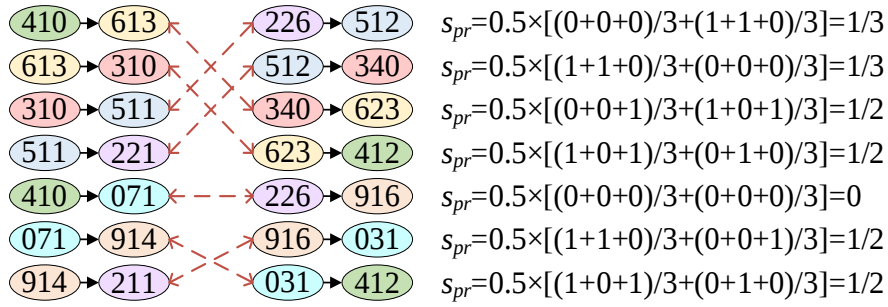


图 3-2 双结构矩阵

3.3.3 共同工序数计算

网状工艺路线间的共同工序可以看成是相似度为 1 的相似工序，所以在上节式(3-13)中已经讨论了共同同工序数的相似情况。在实际中，仍然存在两条网状工艺路线虽然具有共同工序，但这些工序的顺序并不不同的情况。在这种情况下，共同工序的信息将被式(3-15)计算的 s_w 所忽略(如图 2-2 中案例 1 与 2 的区别)。

为了解决这一问题，本文引用了 Paul Jaccard 所提出的 Jaccard 相似系数^[37]，以便独立度量两条网状工艺路线中共同工序的数量：

$$s_j(nos_m, nos_n) = \frac{|\{o_x | o_x \in nos_m \cap o_x \in nos_n\}|}{|\{o_y | o_y \in nos_m \cup o_y \in nos_n\}|} \quad (3-16)$$

在式(3-16)中， o_x 和 o_y 分别表示 nos_m 和 nos_n 两条工艺路线中的其中一个工序。

3.3.4 结构复杂度计算

工艺路线的长度和结构是评估和比较不同工艺路线形状特征的重要指标，长度指的是工艺路线中包含的工序数量，可以通过构建双结构矩阵来量化。结构则描述了工艺路线的组织方式和复杂程度，考虑了路径选择数量、循环和回路等因素。网状工艺路线中往往存在多条可选路径，其结构更为复杂。

本文将网状工艺路线中可选择的工艺路线条数定义为结构复杂度。由此，可以计算两条网状工艺路线结构复杂度的似度：

$$s_s(nos_m, nos_n) = 1 - \frac{\Delta_{pr}}{\max\{|nos_m|, |nos_n|\}} \quad (3-17)$$

在式(3-17)中, Δ_{pr} 是两条网状工艺路线中可选择的工艺路线数之差, $|nos_m|$ 和 $|nos_n|$ 是 nos_m 和 nos_n 这两条网状工艺路线的可选工艺路线数。

3.3.5 多维工艺信息融合的新型网状工艺路线综合相似度

如前文所述, 网状工艺路线中的有效工艺信息包含四类: 工序间相似度、先后顺序关系、共同工序数和结构复杂度, 而本文通过三种中间相似度对其进行了量化, 并表示为 s_w 、 s_j 和 s_s 。在之前的研究中, 大多数采用加权求和的方式得到综合相似度, 而往往忽略了这些中间相似度之间的相关性和内在信息的重复性。图 3-3 清晰地展示了这三种中间相似度之间相互关联的特质。

显而易见, 对于最大匹配权值 s_w 与 Jaccard 相似度 s_j 两者而言, 共同工序的信息也是先后顺序关系的一部分。因为工序的先后顺序关系本身意味着工艺路线应该有相同或者是相似的工序。只有在共同工序顺序完全不同的情况下, 先后顺序关系才无法度量共同工序的相似度。换言之, 工序顺序信息和共同工序信息在先后顺序关系中密切相关。因此, 由图 3-3 可知, s_w 和 s_j 的公共部分是具有相同顺序的共同工序信息, 而 Jaccard 相似度本身可以度量无顺序的共同工序相似度, 最大匹配权重则可以同时度量工序的先后顺序关系和相似工序的相似度。

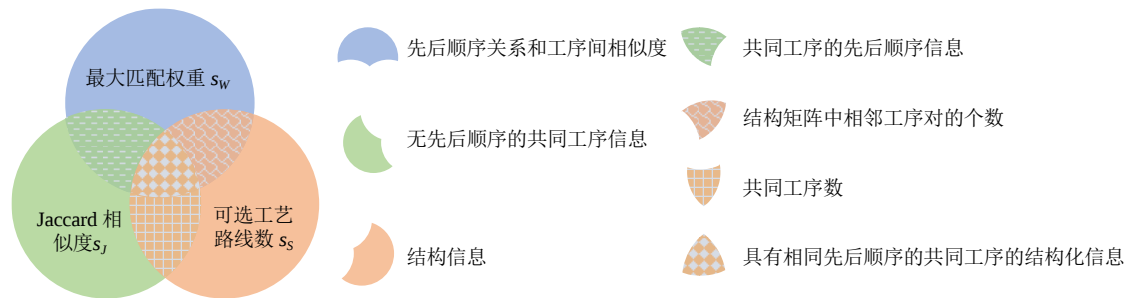


图 3-3 三个中间相似度之间的关系

为了计算出最大匹配权重 s_w 和可选工艺路线数 s_s ，网状工艺路线的结构也必定会影响其工序的先后顺序关系。网状工艺路线的可选工艺路线数越多，对应的双结构矩阵中可以找到的相邻工序对就越多，而该相邻工序对的个数就是 s_w

和 s_s 的共同部分。可选工艺路线数本身可以从整体的角度描述结构信息，而最大匹配权值只考虑了相邻工序对的情况。

对于 Jaccard 相似度 s_j 和可选工艺路线数 s_s 而言，可选工艺路线数越多（即结构复杂度越高），意味着存在共同工序的可能性就越高。因此，这两种度量的共同点是共同工序的数量，而可选工艺路线数和网状工艺路线中所有工序的数量密切相关。

这三种中间相似度的中间部分是具有相同先后顺序的共同工序的结构信息。这意味着这部分信息可以通过这三种方法里任意一种获取得到。为了更直观地解释与分析三种中间相似度的相关性，本节以图 2-2 的 40 条网状工艺路线为例计算了它们的相关系数所得结果如下： $corr(s_j, s_w) = 0.7364$ ， $corr(s_j, s_s) = 0.1295$ 和 $corr(s_w, s_s) = 0.3435$ 。这一结果证明了上述的分析，即这三种度量彼此之间或多或少都存在一定的关联性。

由于这三种中间相似度的相关性，并不适合将它们直接合并为一个加权和，故本文采用 PCA 方法来获得综合相似度，主成分分析的结果如表 3-1 所示。在本文中，选择 y_1 和 y_2 来计算最后的综合相似度值。

表 3-1 三个中间相似度的主成分分析结果

主成分	s_w	s_j	s_s	特征值	比例 (%)
y_1	0.634	0.682	0.365	1.868	62.248
y_2	0.409	0.105	-0.907	0.902	30.062
y_3	0.657	-0.724	0.213	0.231	7.690

在本节中，采用数值仿真的形式来验证本文所提出网状工艺路线综合相似度的有效性，并使用图 2-2 所示的不同相似度情况作为本节的测试集 P 。通过文献[27]中的相关研究，本节采用的性能指标为：

$$R(P, sim) = \frac{\left| \left\{ (nos_m, nos_n) \mid r_{sim}(nos_m, nos_n) = r_r(nos_m, nos_n) \right\} \right|}{\left| \left\{ (nos_m, nos_n) \mid m, n \in [1, N] \right\} \right|} \quad (3-18)$$

其中， sim 是网状工艺路线给定的相似度度量， N 是 P 中工艺路线的条数， $r_{sim}(nos_m, nos_n)$ 是基于 sim 所计算 nos_m 和 nos_n 两条工艺路线的相似度，而

$r_r(nos_m, nos_n)$ 是这两条网状工艺路线的真实相似度。若 $r_{sim} = r_r$ ，则表明通过相似度量 sim 计算所得的相似度与实际情况相符。进而表明，只有当 sim 可以区分两条网状工艺路线之间的微小差异时，这一条件才能成立。因此， R 指标本质是满足 $r_{sim} = r_r$ 的网状工艺路线对个数与测试集 P 中全部工艺路线对的总数之比，且希望 R 指标越大越好。

3.3.6 相似度量性能比较

由于大多数已有研究都集中于线性工艺路线，故将本文所提方法与文献[28]和文献[11]中的相似度量进行比较，其结果如图 3-4 和图 3-5 所示，考虑到文献[28]中体积相似度是预先定义的，在本节中将其视为网状工艺路线中的工序总数目。具体的相似度数値可见附录。附录相似度数値表中，OG 表示原始情况，M 表示最大相似度数値（即 1）。

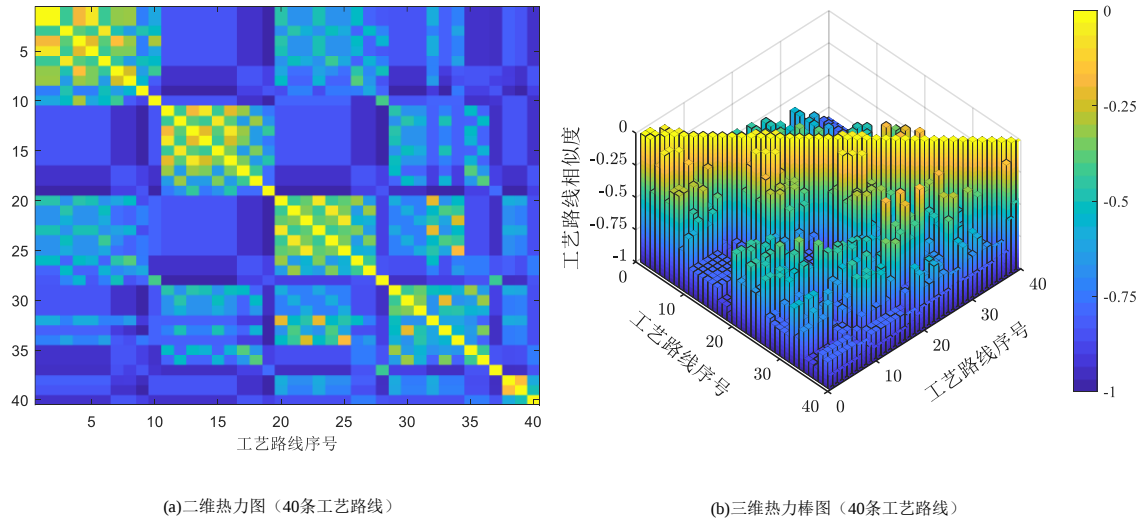


图 3-4 文献[28]设计相似度量计算结果

根据 3.3 节所提相似度计算方法，原本的相似度的范围在 0 到 1 之间，但 AP 算法对于相似度的输入要求是小于 0，所以本文将这两条网状工艺路线集的相似度都转化为 $[-1, 0]$ 。如此，相似度为-1 表示完全不相似，0 表示完全相似。

图 3-4 是测试集 P 根据文献[28]中的相似度量所计算的结果，具体数值可见附录表 A。其中，0 号工艺路线是原始的网状工艺路线（即 OG），其他 39 种

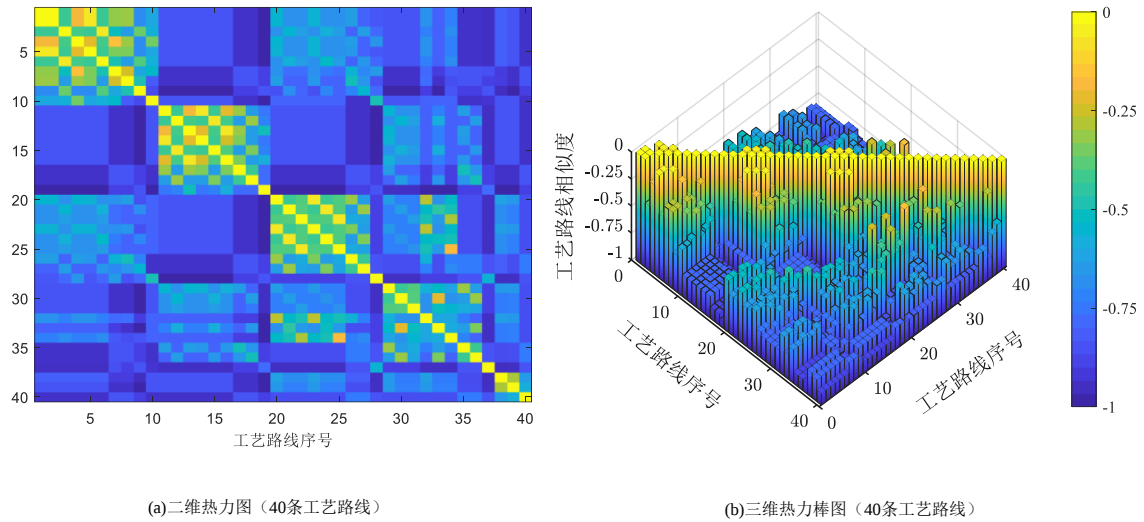


图 3-5 文献[11]设计相似度度量计算结果

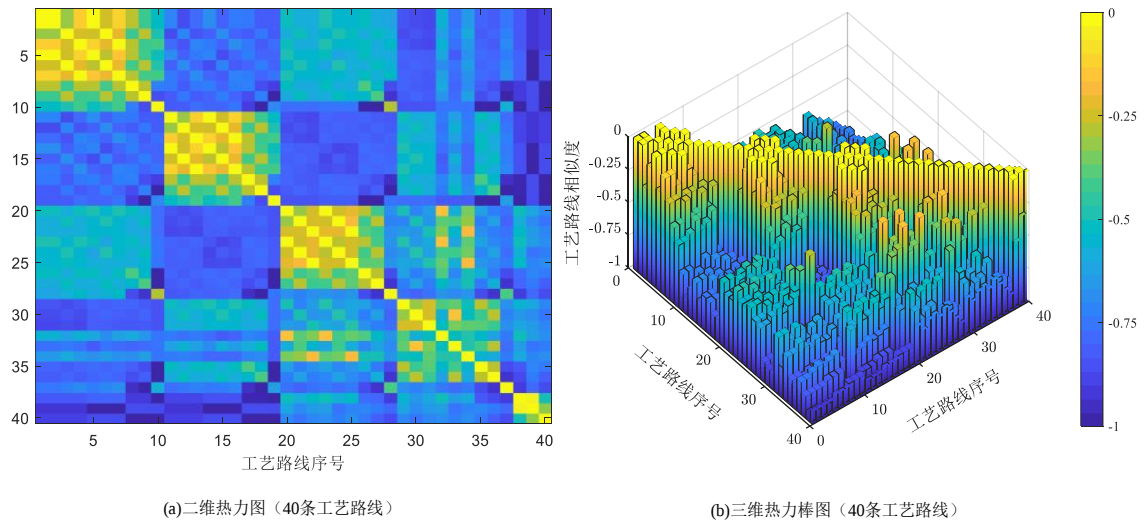


图 3-6 本文设计相似度度量计算结果

相似情况标记为对应数字。在图 3-4 中，明黄色表示为最大相似度数值 0，而深蓝色表示为最小的相似度数值-1。

根据表 2-1 所罗列的网状工艺路线之间不同的相似情况，其他 39 种相似情况与 OG 之间应该具有不同的相似值。若出现两个相似情况具有相同的相似值，则表明该相似度量无法区分这两个相似情况之间的微小差异。如图 3-4 所示，OG 与案例 10、11、12、13、14、15 的相似度值均为 0.2。这反映了文献[28]中提出的相似度量不能有效地度量先后顺序关系与结构复杂性。虽然文献[28]中提出的度量方法可以有效地度量先后顺序关系，但他们只考虑了共同工序信息，没有考虑工序间的相似性，且把相似工序全部视为完全不同的工序。因此，对于文

献[28]所提方法而言，这六种相似案例都与 OG 没有任何共同的工序，属于同种相似情况，故赋予了相同的相似度值。

根据式(3-18)计算可以知道，文献[28]的性能指标为 $R = 19 / 39 = 0.487$ ，此处，19 表示通过该相似度度量可以区分图 2-2 所示的 19 个相似案例。

相较于文献[28]的方法，文献[11]在其基础上额外采用了最长公共子序列来度量网状工艺路线中连续相同工序信息，并结合 Jaccard 相似度来量化相同工序的数量，其所得结果如图 3-5 所示，具体数值可见附录表 C。文献[11]的改进依旧没有考虑工序之间的相似度信息，所以它与文献[28]的计算结果相近，其性能指标也为 $R = 19 / 39 = 0.487$ 。

图 3-6 给出了本文所提多维工艺信息融合的网状工艺路线综合相似度度量的计算结果，具体数值可见附录表 B。显而易见，本文所提方法可以有效区分 36 种不同的相似情况。因此本文方法的性能指标为 $R = 36 / 39 = 0.923$ 。这一结果表明本文所提度量方法相较于文献[28]和文献[11]所提方法而言，考虑和度量的内在工艺信息更多，能更加灵敏地区分网状工艺路线间的细微差异。

本文所设计的相似度度量方法无法区分的相似情况是{17, 18}、{20, 21}和{23, 24}。仔细观察可以看出，图 2-1 已经解释了这些相似情况之间的唯一区别是顺序方面：b)顺序不同，c)部分顺序不同。尽管案例 18、21 和 24 的工序顺序部分不同，但是本文提出的相似度度量方法无法找到这种相同的先后顺序关系。这是因为本文所建立的双结构矩阵要求两个相邻运算之间有严格的连续关系，而无法反映间隔工序之间的先后顺序关系。

图 3-7 表示的是基于本文所提的相似度度量方法计算相似度时 PCA 分析的双标图。可以看出，主成分 y_1 主要描述了隐藏在 s_w 和 s_j 中的信息，而主成分 y_2 主要描述了 s_s 。此外，从图 3-6 可以看出，这些网状工艺路线之间的主要区别在于主成分 y_1 ，因为数据点横向比纵向更分散。

总的来说，从图 3-4，3-5，3-6 三者的计算结果可知，文献[28]和[11]的方法其性能指标都为 0.487，而本文所提相似度度量方法的性能指标为 0.923。这表明对于图 2-2 所示的 39 种不同的相似度情况而言，本文所提方法能有效区分更多的不同情况，即可以更加敏锐地发现不同相似度情况中的共有工艺信息量变化，

从而为不同的情况赋予更加合理和细粒度的相似度值。这一结果验证了本文所提方法在网状工艺路线相似度度量方面的优越性。

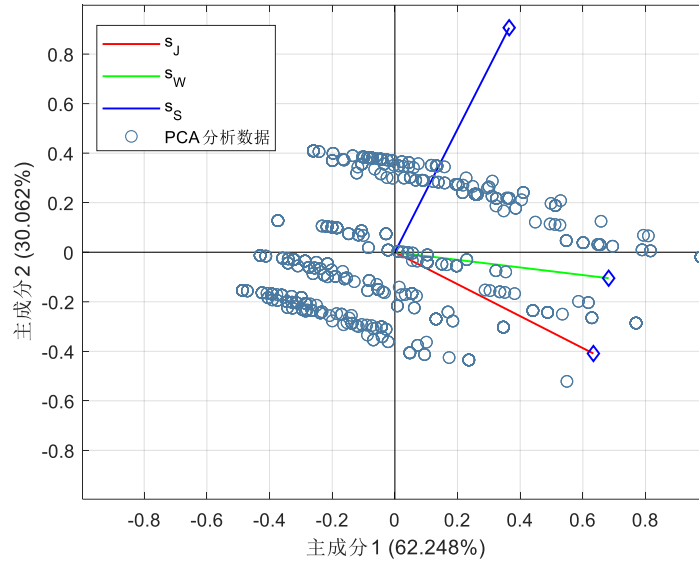


图 3-7 主成分分析双标图

3.4 本章小结

本章提出了一种综合相似度度量方法，以更准确、灵敏地捕捉网状工艺路线之间的细微差异，提升网状工艺路线相似度设计的精确度和实用性。首先，基于国家标准的三级编码方案实现了工序间相似度的量化，便于提取关键工艺信息。其次，通过构建双结构矩阵并结合 KM 算法，将先后顺序关系与工序相似性转化为加权二分图的最大匹配问题，从而增强度量的适用性。此外，引用 Jaccard 系数独立度量无顺序的共同工序相似性，并通过可选工艺路线条数量化结构复杂度，以保证不同工艺路线的结构可比性。基于上述多维信息，通过 PCA 方法去除冗余信息，最终获得更准确的综合相似度。

通过仿真实验验证，本章所提的相似度度量方法在 39 个测试案例中的相似性性能指标达到 0.923，显著优于传统的线性工艺路线度量方法。尽管在部分顺序差异较小的工艺路线中存在难以区分的情况，但整体而言，本章方法相较于传统度量更敏锐地捕捉了网状工艺路线中的共有信息和细微差异，验证了其在工艺路线相似度量化中的优势和创新性。

第 4 章 考虑聚类有效性的典型网状工艺路线发现

4.1 引言

在现代制造领域，复杂产品的生产过程中常常涉及多条网状工艺路线。为提升生产效率和资源利用率，如何从众多网状工艺路线中发现具有代表性的典型网状工艺路线，已成为一项重要的研究课题。传统的聚类算法虽然可以对数据进行分类，但难以满足典型工艺路线发现问题的特殊需求，例如：需要从聚类结果中选择真实存在的工艺路线作为中心、处理复杂的网状结构、容忍噪音数据以及优化聚类有效性。

鉴于上述问题，本章节结合 FHO 算法和 AP 算法，提出一种新型的典型工艺路线发现方法。FHO 作为一种新兴的进化算法，因其全局搜索能力、灵活性和高效性，在参数优化问题中展现了出色的性能；而 AP 算法则以真实数据点作为聚类中心，天然适合发现具有代表性的工艺路线。通过将 FHO 与 AP 结合，本文提出了 FHO-OAP 和 FHO-IAP 两种算法，用于优化参考度和阻尼系数，从而有效提升聚类结果的质量。最后，通过仿真实验验证了该方法在不同数据集上的适用性与优越性。

4.2 所用算法的基础理论

4.2.1 FHO 算法

火鹰优化(Fire Hawk Optimizer, FHO)算法是 2023 年由 Azizi 和 Talatahari 等人^[38]根据黑鸢、啸栗鸢和褐隼的觅食行为提出的进化算法。

FHO 算法模拟了火鹰的觅食行为，包括放火、传火和捕获猎物过程。首先，通过随机初始化确定若干候选解(X)作为火鹰和猎物的位置向量，将它们放置在搜索空间中的初始位置。

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_i^1 & x_i^2 & \cdots & x_i^j & \cdots & x_i^d \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_N^1 & x_N^2 & \cdots & x_N^j & \cdots & x_N^d \end{bmatrix} \quad (4-1)$$

$$x_i^j = x_{i,\min}^j + rand. \cdot (x_{i,\max}^j - x_{i,\min}^j), \quad \begin{cases} i=1,2,\dots,N. \\ j=1,2,\dots,d. \end{cases} \quad (4-2)$$

在公式(4-1)、(4-2)中, N 为搜索空间中候选解的总数; x_i^j 是第 i 个候选解的第 j 个决策变量; $x_i^j(0)$ 表示候选解的初始位置; $x_{i,\min}^j$ 和 $x_{i,\max}^j$ 是第 i 个候选解的第 j 个决策变量的最小和最大边界; $rand$ 是一个[0,1]之间均匀分布的随机数。

为确定火鹰位置, 候选解的目标函数评价考虑了优化问题。火鹰表示较好的解, 其余为猎物。火鹰散布火焰使狩猎更容易, 假设全局最佳解是火鹰首先使用的主火。数学表达式如下:

$$PR = \begin{bmatrix} PR_1 \\ PR_2 \\ \vdots \\ PR_k \\ \vdots \\ PR_m \end{bmatrix}, k=1,2,\dots,m, \quad (4-3)$$

$$FH = \begin{bmatrix} FH_1 \\ FH_2 \\ \vdots \\ FH_l \\ \vdots \\ FH_n \end{bmatrix}, l=1,2,\dots,n, \quad (4-4)$$

其中, PR_k 是搜索空间中关于 m 个猎物总数的第 k 个猎物; FH_l 是搜索空间中 n 个火鹰总数中的第 l 个火鹰。

接下来计算火鹰与猎物之间的总距离, 确定每只火鹰的猎物归属。考虑最近且最优猎物, 处理剩余猎物。 D_k^l 的公式如下:

$$D_k^l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad \begin{cases} l=1,2,\dots,n. \\ k=1,2,\dots,n. \end{cases} \quad (4-5)$$

其中, D_k^l 是第 l 只火鹰与第 k 只猎物之间的总距离; m 是搜索空间中的猎物总数; n 是搜索空间中火鹰的总数。

在算法的下一阶段, 火鹰从主火收集燃烧的木棍, 一边点火一边逼走猎物, 同时寻找其他火源。因此, 这两种行为可作为 FHO 算法主搜索循环中的位置更新过程, 如式(4-6)所示:

$$FH_l^{new} = FH_l + (r_1 \times GB - r_2 \times FH_{Near}), \quad l = 1, 2, \dots, n, \quad (4-6)$$

其中, FH_l^{new} 为第 l 只火鹰(FH_l)的新位置向量; GB 是搜索空间中种群当前最优解, 也就是主火; FH_{Near} 是搜索空间中的另一只火鹰; r_1 和 r_2 是在 (0,1) 范围内的均匀分布的随机数, 用于确定火鹰向主火和其他火影领域的移动。

接下来是猎物位置更新的过程, 如式(4-7)所示:

$$PR_q^{new} = PR_q + (r_3 \times FH_l - r_4 \times SP_l), \quad \begin{cases} l = 1, 2, \dots, n. \\ q = 1, 2, \dots, r. \end{cases} \quad (4-7)$$

其中, PR_q^{new} 为被第 l 只火鹰(FH_l)包围的第 q 个猎物(PR_q)的新位置向量; SP_l 是第 l 只火鹰领域内的一个安全地点; r_3 和 r_4 是在 (0,1) 范围内的均匀分布的随机数, 用于确定猎物向火鹰和安全地点移动。

猎物向其他火鹰领地移动, 可能被伏击或躲避到更安全的地方。在位置更新中考虑这些动作, 使用公式(4-8)表示:

$$PR_q^{new} = PR_q + (r_5 \times FH_{Alter} - r_6 \times SP), \quad \begin{cases} l = 1, 2, \dots, n. \\ q = 1, 2, \dots, r. \end{cases} \quad (4-8)$$

其中, FH_{Alter} 是另一只火鹰; SP 是第 l 只火鹰领地之外的安全地带。

由于自然界中安全场所是大多数动物聚集在一起的场所, 以便在危险期间能够生存, SP_l 和 SP 的表达式如下:

$$SP_l = \frac{\sum_{q=1}^r PR_q}{r}, \quad \begin{cases} q = 1, 2, \dots, r. \\ l = 1, 2, \dots, n. \end{cases} \quad (4-9)$$

$$SP = \frac{\sum_{k=1}^m PR_k}{r}, \quad k = 1, 2, \dots, m. \quad (4-10)$$

具体计算流程如图 4-1 所示:

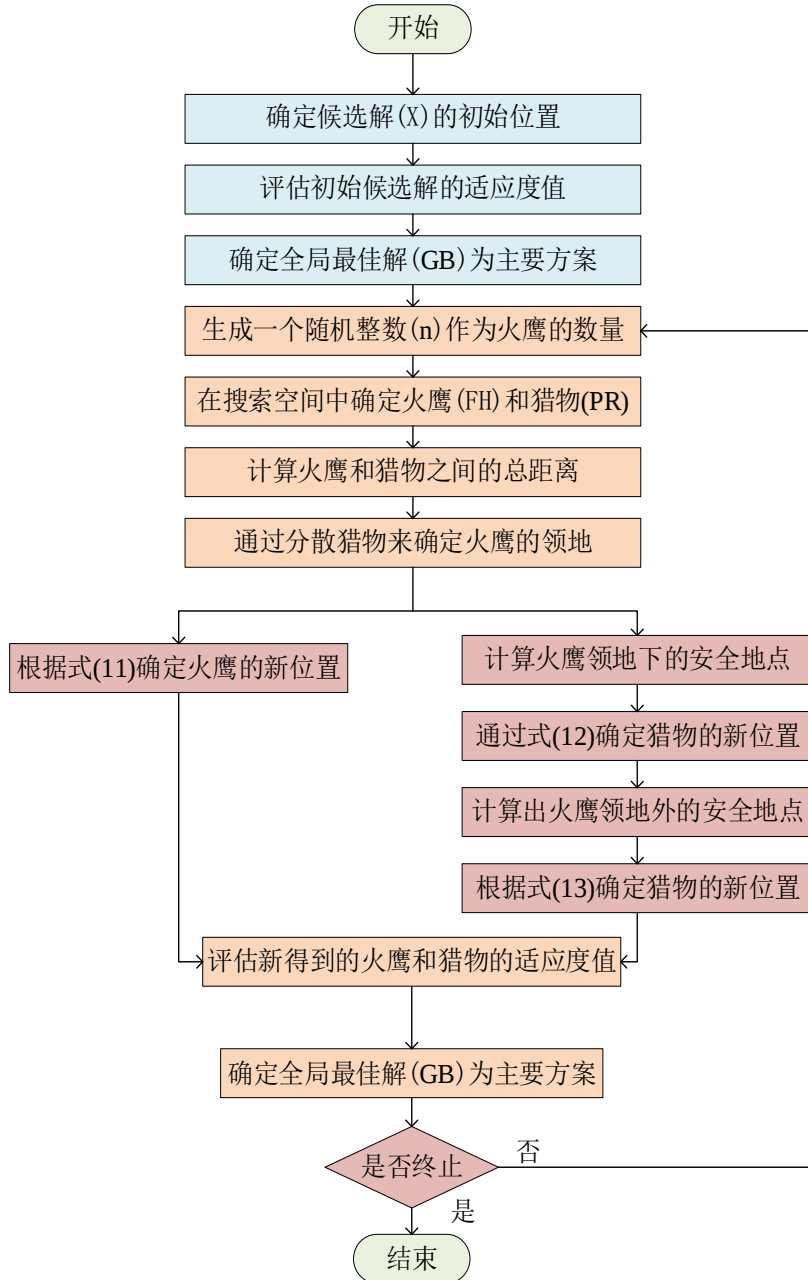


图 4-1 FHO 算法计算流程图

4.2.2 OAP 算法

Frey 和 Dueck 提出了传统的近邻传播算法(Afinity Propagation Clustering, AP)^[52]。该算法是基于数据点间的"信息传递"的一种聚类算法，把一对数据点之间的相似度作为输入，在数据点之间交换真实有价值的信息，直到一个最优的类代表点集合(称为聚类中心或者 exemplar)和聚类逐渐形成。此时所有的数据点到

其最近的类代表点的相似度之和最大。首先，先介绍一下 OAP 算法中的一些相关名词：

(1) 聚类中心(Exemplar): 将所有数据点都作为潜在的聚类中心。

(2) 相似度(Similarity): 数据点 i 和点 j 的相似度记为 $s(i, j)$ ，是指点 j 作为点 i 的聚类中心的相似度。

(3) 参考度(P): 数据点 i 的参考度称为 $p(i)$ 或 $s(i, i)$ ，是指点 i 作为聚类中心的参考度。若按欧氏距离计算其值应为 0，但在 OAP 聚类中其表示数据点 i 作为聚类中心的程度，因此不能为 0。迭代开始前假设所有点成为聚类中心的能力相同，因此参考度一般设为相似度矩阵中所有值得最小值或者中位数，但是参考度越大则说明个数据点成为聚类中心的能力越强，则最终聚类中心的个数越多。

(4) 吸引度(Responsibility): $r(i, k)$ 用来描述点 k 适合作为数据点 i 的聚类中心的程度。从点 i 发送至候选聚类中心点 k ，反映了在考虑其他潜在聚类中心后，点 k 适合作为点 i 的聚类中心的程度。

(5) 归属度(Availability): $a(i, k)$ 用来描述点 i 选择点 k 作为其聚类中心的适合程度。从候选聚类中心点 k 发送至点 i ，反映了在考虑其他点对点 k 成为聚类中心的支持后，点 i 选择点 k 作为聚类中心的合适程度。

(6) 阻尼系数(λ): 为了避免振荡，OAP 算法更新信息时引入了衰减系数 λ 。每条信息被设置为它前次迭代更新值的 λ 倍加上本次信息更新值的 $1-\lambda$ 倍。其中，衰减系数 λ 是介于 0 到 1 之间的实数。

在 OAP 算法中，参考度(P)和阻尼系数(λ)是至关重要的两个参数，它们直接影响着算法的聚类效果和收敛速度。然而，这两个关键参数需要用户手动调节设定，这在一定程度上增加了算法的使用难度和调参成本，需要一定的经验和技巧才能获得理想的结果。

假设数据样本集为 $[x_1, x_2, \dots, x_n]$ ， s 是数据点之间的相似度矩阵，使得 $s(i, j) > s(i, k)$ 当且仅当 x_i 与 x_j 的相似程度大于 x_i 与 x_k 的相似度。AP 算法进行交替两个消息传递的步骤，以更新两个矩阵：

(1) 吸引度矩阵 R : $r(i, k)$ 描述了数据对象 k 适合作为数据对象 i 的聚类中心的程度，表示的是从 i 到 k 的消息。吸引度矩阵的迭代公式为：

$$r_{t+1}(i, k) \leftarrow s(i, k) - \max_{k's.t. k' \neq k} \{a_t(i, k') + s(i, k')\}, \quad (4-11)$$

其中, $a(i, k')$ 表示除 k 外其他点对 i 点的归属度值, 初始值为 0; $s(i, k')$ 表示除 k 外其他点对 i 点的吸引度; $r(i, k)$ 表示数据点 k 成为数据点 i 的聚类中 i 性能的累积证明, $r(i, k)$ 值大于 0, 则表示数据点 k 成为聚类中心的能力强。

(2) 归属度矩阵 A : $a(i, k)$ 描述了数据对象 i 选择数据对象 k 作为其据聚类中心的适合程度, 表示从 k 到 i 的消息。归属度矩阵的迭代公式为:

$$a_{t+1}(i, k) = \begin{cases} \min \left\{ 0, r_t(k, k) + \sum_{i's.t. i' \notin \{i, k\}} \max \{0, r_t(i', k)\} \right\}, & i \neq k, \\ \sum_{i's.t. i' \notin \{i, k\}} \max \{0, r_t(i', k)\}, & i = k \end{cases} \quad (4-12)$$

其中, $r(i', k)$ 表示点 k 作为除 i 外其他点的聚类中心的相似度值, 取所有大于等于 0 的吸引度值, 加上 k 作为聚类中心的可能性。即点 k 在这些吸引度值大于 0 的数据点的支持下, 数据点 i 选择 k 作为其聚类中心的累积证明。

为了优化吸引度和归属度, 算法将进行多次迭代。当迭代过程达到以下任一条件时, 算法将终止: 一是经过多次迭代后, 所有决策变量保持不变, 说明算法已收敛至最优解; 二是算法执行次数超过预设的迭代上限, 防止算法陷入无限循环; 三是在某个小区域内, 样本点的决策在多次迭代后稳定不变, 表明该区域已达到局部最优状态。此时, 算法将结束迭代过程, 输出最终结果。

为了避免振荡, OAP 算法更新信息时引入了衰减系数 λ 。每条信息被设置为它前次迭代更新值的 λ 倍加上本次信息更新值的 $1-\lambda$ 倍。其中, 衰减系数 λ 是介于 0 到 1 之间的实数, 一般取 0.5。即第 $t+1$ 次 $r(i, k)$ 、 $a(i, k)$ 的迭代值:

$$r_{t+1}(i, k) \leftarrow (1-\lambda) * r_{t+1}(i, k) + \lambda * r_{t+1}(i, k) \quad (4-13)$$

$$a_{t+1}(i, k) \leftarrow (1-\lambda) * a_{t+1}(i, k) + \lambda * a_{t+1}(i, k) \quad (4-14)$$

Givoni 和 Frey^[47]指出 OAP 算法本质上是一种基于因子图的最大和置信传播算法。定义了二元隐式变量集 $C = \{c_{11}, c_{12}, \dots, c_{MN}\}$ 为因子图中的变量和聚类结果的分配矩阵。OAP 算法因子图如图 4-2 所示:

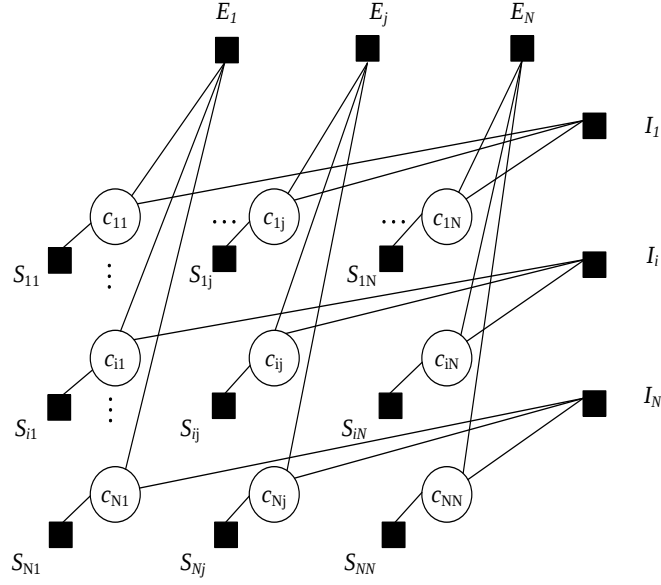


图 4-2 OAP 算法因子图

4.2.3 IAP 算法

对于一个含有 N 个工艺路线的集合 $X = \{x_1, \dots, x_N\}$, 定义 $S = \{sim(x_i, x_j)\}$ 为 N 中所有工艺路线之间的相似度集。文献[35]在 OAP 算法中加入了数量软约束和半径软约束, 定义 $q_e \geq 0$ 和 $q_r \geq 0$ 分别是数量软约束和半径软约束的罚值。IAP 算法因子图如图 4-3 所示:

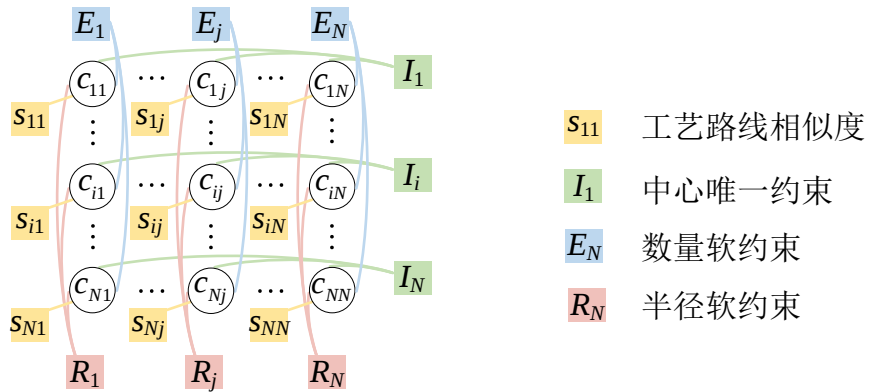


图 4-3 IAP 算法因子图

在 IAP 算法中一共有四种约束: 中心唯一约束、中心一致约束, 数量软约束与半径软约束。中心唯一约束是指每个数据点有且只有一个中心。中心一致约束

是指只有某个数据点选择自己作为中心的时候，其他数据点才可以选择该点作为它们的中心。在本文中，中心一致约束也包含在数量软约束和半径软约束中。其中， I 、 E 和 R 分别表示中心唯一约束、数量约束和半径约束。用公式表示为：

$$I_i(c_{i1}, \dots, c_{iN}) = \begin{cases} -\infty & \sum_{j=1}^N c_{ij} \neq 1 \\ 0 & \text{其他} \end{cases} \quad (4-15)$$

$$E_j(c_{1j}, \dots, c_{Nj}) = \begin{cases} -\infty & c_{jj} = 0, \exists c_{ij} = 1, i \neq j \\ \min \left[-\left(n_{clstr} - \sum_{i=1}^N c_{ij} \right) q_e, 0 \right] & c_{jj} = 1 \\ 0 & \text{其他} \end{cases} \quad (4-16)$$

$$R_j(c_{1j}, \dots, c_{Nj}) = \begin{cases} -\infty & c_{jj} = 0, \exists c_{ij} = 1, i \neq j \\ \sum_{\{i|c_{ij}=1, i \neq j\}} \min \left[-\frac{\varepsilon_{clstr} - \text{sim}(x_i, x_j)}{|\varepsilon_{clstr}|} q_r, 0 \right] & c_{jj} = 1 \\ 0 & \text{其他} \end{cases} \quad (4-17)$$

由上式可知，当 $q_e = 0$ 或者 $q_r = 0$ 时， E 和 R 就会简化为中心一致约束。

S 表示为 x_i 和 x_j 之间的相似度：

$$S_{ij}(c_{ij}) = \begin{cases} \text{sim}(x_i, x_j) & c_{ij} = 1 \\ 0 & c_{ij} = 0 \end{cases} \quad (4-18)$$

因此，IAP 算法的目标函数为：

$$\begin{aligned} netsim = \arg \max_C & \sum_{i=1}^N S_{ij}(c_{ij}) + \sum_{i=1}^N I_i(c_{i1}, \dots, c_{iN}) \\ & + \sum_{j=1}^N E_j(c_{1j}, \dots, c_{Nj}) + \sum_{j=1}^N R_j(c_{1j}, \dots, c_{Nj}) \end{aligned} \quad (4-19)$$

图 4-4 展示了在 IAP 算法因子图中，功能节点与变量节点间进行信息交换的流程。图 4-4 标出 IAP 算法中的七类信息传递机制：其中， α 与 β 代表在变量节点与功能节点 E 间的信息传递； η 与 β 则对应于变量节点与功能节点 I 之间的信息传递； λ 与 τ 分别描述了变量节点与功能节点 R 的信息传递过程；而 $\text{sim}(x_i, x_j)$ 则体现了功能节点 S 向变量节点信息传递。作为一种采用最大和信息更新公式的最大和置信传播算法，这些信息的更新均遵循特定的规则进行：

$$\mu_{v \rightarrow f}(x) = \sum_{\{i | f_i \in ne(v) \setminus f\}} \mu_{f_i \rightarrow v}(v) \quad (4-20)$$

$$\mu_{f \rightarrow v}(v) = \max_{v_1, \dots, v_N} \left[f(v_1, \dots, v_N) + \sum_{\{j | x_j \in ne(f) \setminus v\}} \mu_{v_j \rightarrow f}(v_j) \right] \quad (4-21)$$

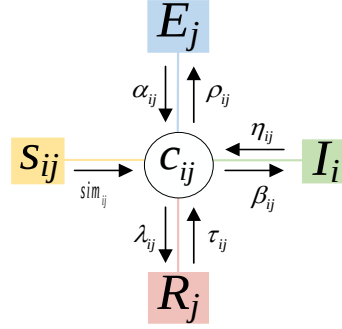


图 4-4 因子图细节

其中，变量节点到功能节点的信息表示为 $\mu_{v \rightarrow f}$ ，功能节点到变量节点的信息表示为 $\mu_{f \rightarrow v}$ ，变量节点 v 除功能节点 f 外的邻节点集合表示为 $ne(v) \setminus f$ ，功能节点 f 除变量节点 v 外的邻节点集合表示为 $ne(f) \setminus v$ 。

(1) 从变量节点到功能节点的信息更新

对于功能节点 E_j ，更新公式为：

$$\rho_{ij} = \rho_{ij}(1) - \rho_{ij}(0) = sim(x_i, x_j) + \tau_{ij} + \eta_{ij} \quad (4-22)$$

其中， $\rho_{ij}(1)$ 和 $\rho_{ij}(0)$ 分别为 $c_{ij} = 1$ 和 $c_{ij} = 0$ 时节点 c_{ij} 传给节点 E_j 的信息。

对于功能节点 I_i 和 R_j ，更新公式为：

$$\beta_{ij} = sim(x_i + x_j) + \tau_{ij} + \alpha_{ij} \quad (4-23)$$

$$\lambda_{ij} = sim(x_i + x_j) + \eta_{ij} + \alpha_{ij} \quad (4-24)$$

(2) 从功能节点到变量节点的信息更新

对于功能节点 S_{ij} ，其更新公式可由式(4-18)和式(4-21)推导所得：

$$\begin{aligned} sim_{ij} &= sim_{ij}(1) - sim_{ij}(0) \\ &= \max_{c_{ij}=1} [S_{ij}(1) + 0] - \max_{c_{ij}=0} [S_{ij}(0) + 0] \\ &= sim(x_i, x_j) \end{aligned} \quad (4-25)$$

4.3 基于 FHO 优化 AP 的典型工艺路线发现方法

4.3.1 总体思路与框架

聚类算法是一种无监督学习方法，用于将数据集中的对象分成不同的组，使得同一组内的对象具有相似的特征。传统聚类算法包括 K-medoids 算法^[48]、层次聚类算法^[49]、密度聚类算法^[50]和谱聚类算法^[51]等。

不同于传统聚类问题旨在将数据合理聚类，从而划分成不同的簇来发现它们的共性。典型网状工艺路线的发现问题具有以下特点：

(1) 不仅需要对已有的网状工艺路线聚类，还要从聚类簇中选择一条真实存在的网状工艺路线作为聚类中心，并将其视为该聚类簇的典型工艺路线。

(2) 由于网状工艺路线结构复杂，网状工艺路线之间存在多种可能的关联和依赖关系。这就导致网状工艺路线数据集的数据分布不规则且难以预估。

(3) 网状工艺路线的聚类过程需要考虑聚类有效性问题，且更加关注发现的典型工艺路线是否具有代表性，而不是关注簇间距离或者簇内总相似度。

(4) 网状工艺路线中可能会存在噪音数据。

相比于传统聚类方法，AP 算法通过基于相似度的消息传递机制来自动确定数据的最佳聚类数量，对复杂的非线性数据集和噪音数据有一定的容忍性，具有较高的稳定性。另一方面，AP 算法最显著的特点是将已有的真实数据点作为聚类中心，而不是生成一个虚拟的聚类簇中心。因此，AP 算法非常适合于典型工艺路线的发现问题的发现，而它所得聚类结果中各聚类簇的聚类中心就是本文所发现的典型网状工艺路线。

相关研究已表明^[52]，AP 算法的参考度(P)和阻尼系数(λ)会直接影响算法最终的聚类结果。而对于考虑了聚类有效性的典型工艺路线发现问题而言，想要人为设置最优的参数显然是不可能的。为此，本文考虑通过智能优化算法来对这两个参数进行自动调优，从而保证 AP 算法最终聚类结果的质量。

FHO 算法是 2023 年由 Mahdi 等人^[38]根据火鹰的觅食行为提出的进化算法，该算法具有高效性、灵活性、简单性、全局搜索能力和收敛性等优点。相关实验

已经证明，它在处理复杂问题时相较于其他已有算法效果更好，性能更高，故适合用于为 AP 算法自动选择合适的优化参数。

在本文中，将 FHO 与 OAP^[52]和 IAP^[35]相结合，用于求解考虑聚类有效性的典型网状工艺路线发现问题。具体而言，FHO 算法首先从参考度(P)和阻尼系数(λ)的取值范围中随机选取数值，并将其与网状工艺路线集合的相似度矩阵一起作为 AP 算法的输入。AP 算法基于输入得到聚类结果，并根据式(2-1)计算 *netsim*。FHO 算法将计算所得 *netsim* 作为对应参数的适应度值，再基于此进行迭代更新，并最终找到可以得到最佳 *netsim* 的参考度和阻尼系数，同时输出最终的聚类结果。

4.3.2 基于 FHO-OAP 的典型工艺路线发现方法

FHO-OAP 算法相较于 OAP 算法有两大改进：一是采用 FHO 算法自动调节选择最优的 P 与 λ ；二是由于 OAP 算法在聚类过程中并没有考虑数量软约束和半径软约束，故在 OAP 算法对网状工艺路线数据集进行聚类得到聚类结果之后，需要再根据式(2-1)所计算数量软约束与半径软约束的罚值，并进行累加。最终累加得到的最终结果记为式(2-1)中的 *netsim*，并视为 FHO 中个体的适应度值。*netsim* 越大，也就是所有网状工艺路线到其对应中心的相似度之和加上各软约束罚值的累加越大，表示对应聚类结果的效果越好。FHO-OAP 具体计算流程如图 4-5 所示。在图 4-5 中，蓝色框图表示的是 OAP 算法部分，黄色的框图表示的是 FHO 算法部分。FHO-OAP 算法如算法 4.1 所示。

在 FHO-OAP 算法中，初始化 FHO 种群就是在两个参数的有效范围中，随机选取初始的取值。火鹰 FH 与猎物 PR 是初始解经过 OAP 算法计算评估后，较好的 P 与 λ 表示为火鹰，其余为猎物，而 GB 则是目标函数所求的最大 *netsim*。在图 4-5 中，评估更新后的火鹰和猎物的适合度值，就是比较新得到的 *netsim* 是否比 GB 大。若 GB 不是全局最优解，就表明聚类效果不是最好的一次，则继续在搜索空间中确定新的 p 与 λ 继续迭代寻找最大的 *netsim*。当 *netsim* 最大时，就将最大的 *netsim* 赋值给 GB。此时，算法所得结果就是考虑聚类有效性的最佳聚类效果，而每个聚类簇的聚类中心就是本文所求的典型网状工艺路线。

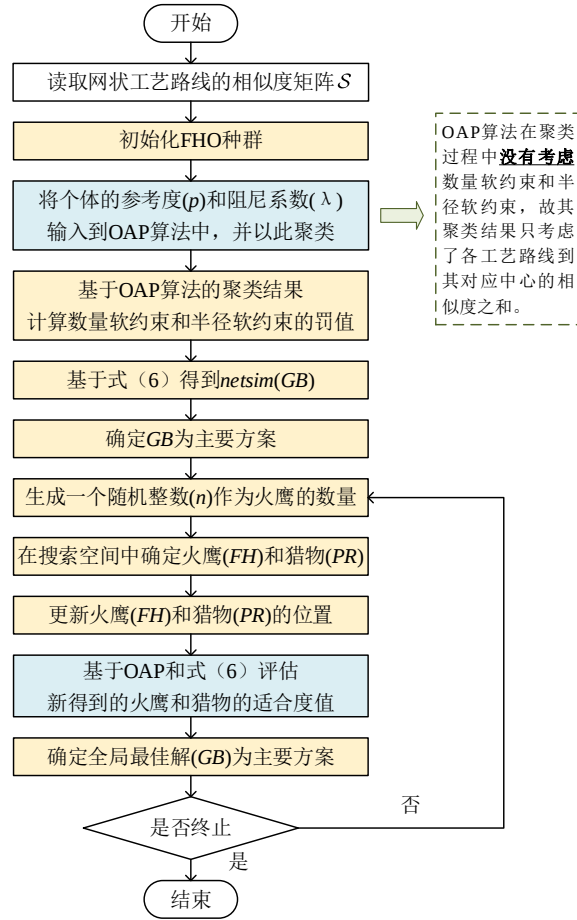


图 4-5 FHO-OAP 算法流程图

算法 4.1 FHO-OAP 算法

1. 读取相似度数值
2. 设定 p 与 λ 预估范围, 将式(1)设为目标函数
3. 设置最大迭代次数, 选定候选解初始个数, 随机生成火鹰数量
4. 通过 AP 算法计算出初始聚类结果
5. 根据式(1)在初始聚类结果中加入两个软约束, 计算初始候选解的目标函数值 $netsim$
6. 选定初始最优目标函数值为 GB
7. 在搜索空间中确定火鹰(FH)和猎物(PR)的位置
8. 计算 FH 与 PR 之间的距离
9. **for** $i = 1 : FH$ **do**
10. 计算 FH 与 PR 之间的距离
11. **end**
12. 更新种群位置
13. **while** 函数评估次数 < 最大迭代次数
14. $it = it + 1$
15. 更新 FH 位置

-
16. 更新 PR 位置
 17. 评估新得到的 FH 和 PR 的目标函数值
 18. 比较当前最优 $netsim$ 与 GB
 19. 将最优 $netsim$ 赋予 GB
 20. **end**
 21. 输出最优聚类结果
-

4.3.3 基于 FHO-IAP 的典型工艺路线发现方法

OAP 在聚类网状工艺路线时, 着重于工艺路线之间的相似度, 而忽略了聚类有效性的考量。具体来说, OAP 本身并不会直接考虑 2.5 节中提到的数量软约束和半径软约束。因此, 在 OAP 完成初始聚类后, FHO-OAP 会进一步介入, 它会基于 OAP 的聚类结果, 计算违反数量软约束和半径软约束所产生的罚值。这些罚值将被用于评估聚类结果的有效性, 并最终得到相应的 $netsim$, 从而指导进一步的优化, 旨在找到同时兼顾工艺路线相似度和聚类有效性的更优聚类方案。

与 OAP 不同, 文献[35]提出的 IAP 直接将数量软约束和半径软约束融入到 AP 算法的聚类过程中。为此, IAP 在迭代过程中需要同步更新四种类型的信息变量, 以反映软约束的影响。这种做法的优势在于, IAP 在聚类时不仅考虑了工艺路线之间的相似度, 还兼顾了两个软约束, 从而力求获得一个使 $netsim$ 值尽可能大的折衷方案, 实现相似度和聚类有效性的平衡。然而, 这种集成的代价是 IAP 的聚类结果质量对参考度(P)和阻尼系数(λ)这两个关键参数的依赖性更高。不合适的参数设置可能会导致算法难以收敛或陷入局部最优, 最终影响聚类效果。因此, 在本节中, 同样将 FHO 用于优化 IAP, 从而使其能自动获得使 $netsim$ 最大的相应 P 值和 λ 值。

由图 4-5 和图 4-6 可知, FHO-IAP 算法与 FHO-OAP 算法的最大区别在于软约束罚值的加入位置。具体而言, FHO-OAP 算法在 AP 算法对工艺路线数据库进行聚类得到最终的聚类结果之后, 就加入了两个软约束的罚值。这样的不同带来的影响在于, FHO-OAP 中 OAP 的聚类过程更加简单, 因为它没有考虑聚类有效性。相反, FHO-IAP 中 IAP 的聚类过程则要同时考虑软约束带来的负面影响。因此, 引入 FHO 对 IAP 所带来的性能提升会更加明显。

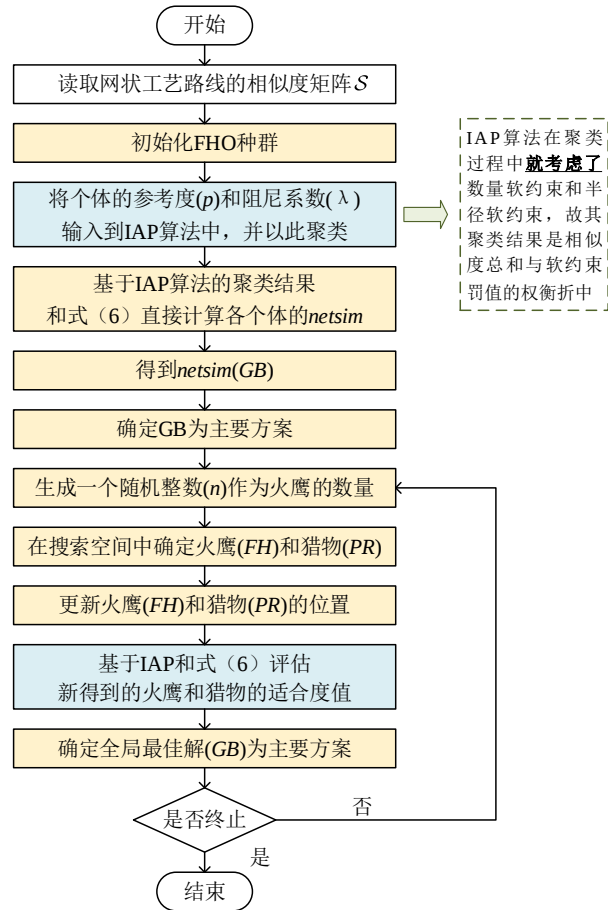


图 4-6 FHO-IAP 算法流程图

算法 4.2 FHO-IAP 算法

1. 读取相似度数值
2. 设定 p 与 λ 预估范围
3. 将聚类过程中加入罚值的式(1)设为目标函数
4. 设置最大迭代次数, 选定候选解初始个数, 随机生成火鹰数量
5. 计算初始候选解的目标函数值 $netsim$, 选定初始最优目标函数值为 GB
6. 在搜索空间中确定火鹰(FH)和猎物(PR)的位置
7. 计算 FH 与 PR 之间的距离
8. **for** $i = 1 : FH$ **do**
9. 计算 FH 与 PR 之间的距离
10. **end**
11. 更新种群位置
12. **while** 函数评估次数<最大迭代次数
13. $it=it+1$
14. 更新 FH 位置
15. 更新 PR 位置
16. 评估新得到的 FH 和 PR 的目标函数值
17. 比较当前最优 $netsim$ 与 GB

-
18. 将最优 *netsim* 赋予 *GB*
 19. **end**
 20. 输出最优聚类结果
-

4.4 仿真实验验证

由于网状工艺路线数据集中数据分布存在显著的未知性与复杂性,涵盖数据密度不均、离散程度各异等特征,因此典型网状工艺路线的发现方法必须具备良好的鲁棒性与适应性,以确保能够有效处理不同数据分布特征的工艺路线集。为了全面分析所提基于 FHO 优化 AP 的典型工艺路线发现方法的聚类效果和优化性能,本文基于两种不同的网状工艺路线数据集来测试所提出算法的性能,一个是图 2-2 所示 40 条网状工艺路线组成的数据集,一个是随机生成的包含 100 条网状工艺路线的数据集。同时,为了有效分析和验证本文所提方法的有效性,将其与已有的 OAP 算法(2007)^[52]、平均链接聚类(Average Linkage Clustering, AL C)算法(2016)^[18, 53]、IAP 算法(2019)^[35]、改进 K-medoids(Improve K-medoids, IK M)算法(2021)^[27]进行了对比分析,这些方法都是在典型工艺路线发现领域常用的方法。

4.4.1 聚类性能指标

传统聚类分析过程侧重于将数据合理分类,强调类间距离尽可能大而类内距离尽可能小^[40]。然而,与传统聚类问题不同,本文研究的典型网状工艺路线发现问题旨在通过聚类过程从工艺路线数据集中筛选出最具代表性的典型工艺路线,以最大限度地反映聚类簇中其他数据所包含的共有特性。为了合理评估聚类结果的有效性,本文选择式(4-19)所示的 *netsim* 作为本节的性能指标,该指标本质上是各工艺路线与其对应中心的相似度值以及两个软约束罚值的累加。

4.4.2 仿真实验设置

本节中对于仿真实验参数设置如下: $n = \text{round}(N / K)$, n 表示数量软约束下有效网状工艺路线聚类簇中工艺路线的最小数量; $\varepsilon = \min s(x_i, x_j) / K$, ε 为半径软约束下有效网状工艺路线聚类簇的最大半径; $q_e = 0.5$ 、 $q_r = 1$, q_e 和 q_r 分别是数

量软约束和半径软约束的罚值； K 表示不同情景下网状工艺路线数据集的理想聚类个数。

在本节的仿真实验中，主要是通过理想聚类个数 K 来设置不同的数量软约束 n 和半径软约束 ε ，以模拟不同情景和需求下的典型工艺路线发现问题。 q_e 和 q_r 是根据相似度范围和两个软约束的罚值来设计计算的。具体地说，由式(2-1)可知，数量软约束的罚值直接由违反数量软约束的工艺路线数量决定，而半径软约束的罚值则由各工艺路线与其中心的相似度值和 ε 之间的比值所决定的。

本节中，设置 K 的取值范围为 1 到 8，用来表示 8 个不同情景下的典型工艺路线发现问题。这 8 个问题虽然使用的是相同的数据集，然而，在聚类分析中，数量软约束侧重于对聚类样本数量的弹性限制，而半径软约束则聚焦于聚类空间范围的动态界定。由于二者在聚类过程中考虑的侧重点存在显著差异，因此它们对聚类过程的影响机制以及最终形成的聚类结果均会产生明显区别。算法需要在各种不同的实际应用场景中进行全面且系统的性能测试，以充分展示其良好的适用性、鲁棒性以及在不同条件下的表现。在每次确定 K 的数值之后，各个算法重复运行 5 次并取均值和方差来衡量算法的真实性能。换句话说，为了确保结果的可靠性，每个算法在每个数据集上都需要进行 8×5 次独立的重复运行。

本文参考度(P)的搜索空间根据文献[53]知， P 的下限为 $p = dpsim1 - dpsim2$ ，具体见公式(4-26)、(4-27)。根据文献[54]知 P 的上限取值 $pm/2$ 。阻尼系数 λ 的搜索范围为[0,0.99]。

$$dpsim1 = \max \left\{ \sum s(i, j) \right\} \quad (4-26)$$

$$dpsim2 = \max_{i=j} \left\{ \sum_k \max \{ s(i, k), s(j, k) \} \right\} \quad (4-27)$$

4.4.3 基于小型数据集的性能分析

本节以图 2-2 所示的 40 条网状工艺路线作为测试数据集，其相似度分布如图 3-6 所示。表 4-1 给出了六种算法在该网状工艺路线数据集上运行所得的聚类结果均值与方差。其中，加粗字体表示了相同问题配置下各算法所得的最优结果。同时，基于 Friedman 检验^[55]计算了各算法的平均排名(Rank)。

表 4-1 40 条网状工艺路线聚类结果与方差

K	OAP	ALC	IAP	IKM	FHO-OAP	FHO-IAP
1	-22.033(0.00)	-22.033(0.00)	-22.033(0.00)	-22.033(0.00)	-22.033(0.00)	-22.032(0.00)
2	-20.941(0.00)	-38.308(0.00)	-21.201(0.00)	-19.738(0.00)	-20.941(0.00)	-19.736(0.00)
3	-21.627(0.00)	-32.483(0.00)	-21.627(0.00)	-21.627(0.00)	-21.627(0.00)	-21.627(0.00)
4	-22.892(0.00)	-40.839(0.00)	-22.830(0.21)	-22.962(1.58)	-22.892(0.00)	-22.444(0.00)
5	-24.575(0.00)	-41.292(0.00)	-24.616(0.01)	-24.575(0.00)	-24.575(0.00)	-24.575(0.00)
6	-28.959(0.00)	-29.684(0.00)	-28.458(0.00)	-29.206(0.35)	-28.459(0.00)	-28.456(0.00)
7	-29.945(0.00)	-35.309(0.00)	-30.266(0.00)	-31.277(0.27)	-30.140(0.43)	-30.099(0.027)
8	-31.291(0.00)	-40.518(0.00)	-35.679(0.00)	-32.467(0.78)	-30.061(0.00)	-30.192(0.035)
Rank	3.0625	5.7500	3.7500	3.8125	2.9375	1.6875

从表 4-1 可知，本文所提 FHO-IAP 和 FHO-OAP 的聚类效果明显好于已有的其他四种聚类算法，这是因为 FHO-IAP 和 FHO-OAP 不仅在聚类过程中就考虑了基于聚类有效性的数量软约束和半径软约束，且通过 FHO 算法有效地优化了 OAP 算法和 IAP 算法中的 p 和 λ ，提高了原本 OAP 和 IAP 算法的整体性能。

对比 OAP 算法和 IAP 算法可以发现，OAP 算法反而聚类效果更好。可能的原因是本节所用数据集只有 40 条数据，求解空间相对较小，OAP 很容易找到最好的聚类结果，而相同算法参数下。IAP 需要在聚类过程中考虑两个软约束，求解空间更大，求解过程也更复杂，更加不易找到最好的结果。

另一方面，对比 OAP 算法和 FHO-OAP 算法可知这两个算法的聚类效果差距不大；但 FHO-IAP 算法的聚类效果相较于 IAP 算法而言明显进步很多。这是因为 OAP 算法本身更新传递的中间信息较少，计算过程相对简单，所以即使加入 FHO 算法来优化参数 OAP 提升的聚类效果也微乎其微。而 IAP 算法在聚类过程中需要考虑两个软约束的情况，即需要从聚类结果和软约束罚值之间找到最佳的平衡点，所以考虑和更新的中间信息较多，问题更复杂。在这种情况下，加入 FHO 算法优化参数之后，其聚类效果提升更为明显。

最后，对于 FHO-OAP 算法而言，FHO-IAP 算法聚类效果更好。这是因为 FHO-IAP 算法是将两个软约束加入到了 IAP 算法的聚类过程当中，保证了 FHO-IAP 的最后结果是同时考虑了工艺路线的聚类情况和两个软约束带来的罚值，从

而从根本上保证其最终的目的是使得式(4-19)最大。而 FHO-OAP 算法则是在 OAP 聚类结束之后直接将两个约束的罚值加入到聚类结果当中，这样导致软约束并未直接参与聚类，所以 FHO-OAP 算法没有 FHO-IAP 算法聚类效果好。

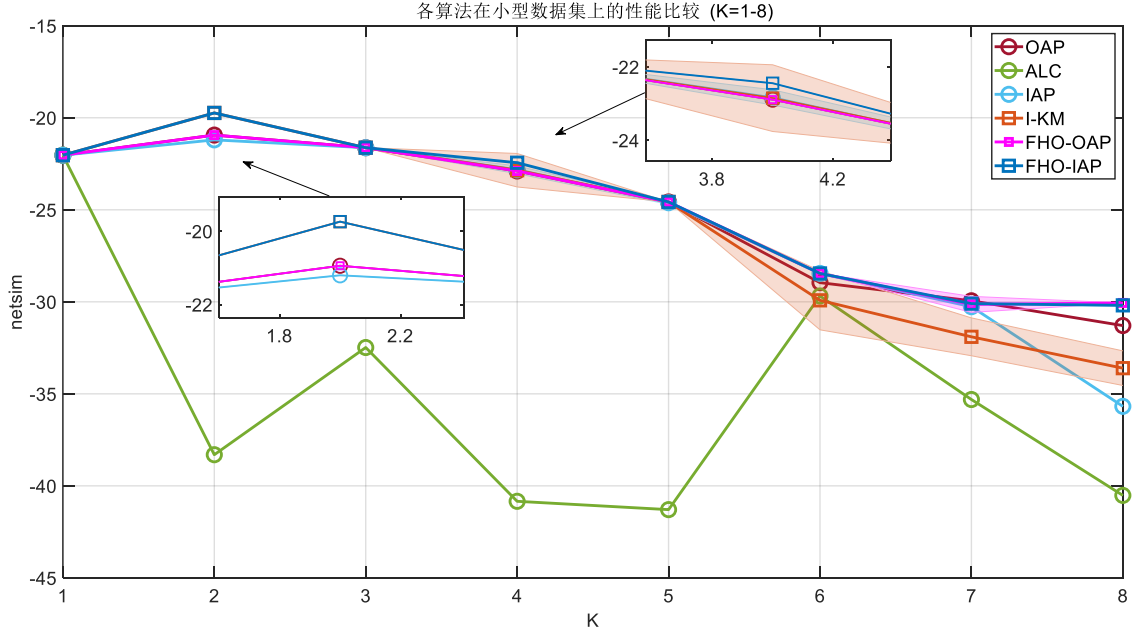


图 4-7 6 种算法对 40 条网状工艺路线聚类效果对比图

图 4-7 显示了六种聚类算法在小型数据集中随值变化的聚类效果变化趋势，图中彩色面积表示的是各算法计算出 *netsim* 的方差（即表 4-1 中括号内的数值）方差是衡量算法结果离散程度的重要统计量，它通过计算每次算法结果与其均值之间差异的平方，反映了数据的波动性。方差的公式如下所示：

$$\sigma^2 = \frac{1}{5} \sum_{i=1}^5 (cr_i - \mu)^2 \quad (4-28)$$

在式(4-28)中， σ^2 即为上述方差，它是对各算法所得 *netsim* 离散程度的一种度量。该数值越大，表示 *netsim* 的分散程度越高，图 4-3 中的彩色面积就越大，意味着对应算法的聚类结果越不稳定。 cr_i 是算法运行第 i 次的聚类结果，而 μ 则表示算法重复运行 5 次所得聚类结果的平均值。

图 4-7 表明，当 $K \leq 5$ 时，IKM 算法、FHO-IAP 算法和 FHO-OAP 算法的聚类效果在这六种聚类算法中性能较好，但随着 K 的增加，这三个算法的性能差距逐渐增大。具体而言，由于 IKM 对初始解的依赖严重，当 K 逐渐增大时，初始

解的随机性变大,故所得结果的方差变大,图中彩色部分面积变大,说明了 IKM 算法在 K 值较大时聚类效果极不稳定。相较而言, FHO-OAP 算法和 FHO-IAP 算法在 K 值较大的时候依旧保持了较好的结果和较高的稳健性。

4.4.4 基于大型数据集的性能分析

在本节中,采用一个随机生成的包含 100 条网状工艺路线的数据集来验证本文所提的算法性能,在该数据集中,各工序采用三位编码表示,随机生成过程服从 001 到 999 的离散均匀分布。每条网状工艺路线的长度服从离散均匀分布,且长度最大为 8。图 4-8 是 100 条网状工艺路线集的相似度分布情况,不同于图 3-6,由于本文所提综合相似度考虑的工艺信息较多,相似度计算过程更加复杂,故随机生成的数据集其相似度结果呈现出明显的两极分化。

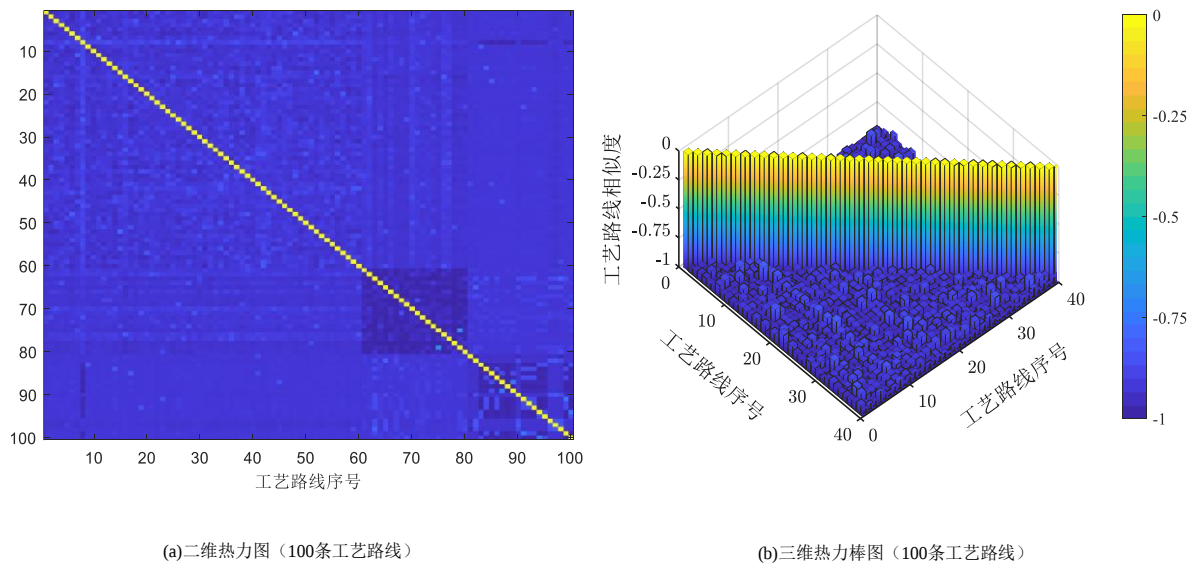


图 4-8 100 网状条工艺路线集相似度分布情况

表 4-2 给出了这六种算法对于 100 条网状工艺路线集的聚类结果均值与方差,并运用加粗字体来表示值相同情况下,所有算法里最优的那个聚类结果。此外同样基于 Friedman 检验^[55]计算了各算法在大数据集上的性能平均排名(Rank)。

从表 4-2 中可以看出, FHO-IAP 算法是该六种算法中聚类效果最好的,并且因为将算法应用在大数据集中与其他五种算法的聚类结果相比,所得聚类结果优异显著。这表明 FHO-IAP 的算法在不同大小和数据分布的数据集上都有较好的效果,具有较强的普适性。

此外,从表 4-2 也可以看出,不同于小数据集的结果,所有在聚类过程中就考虑软约束的算法(包括 IAP、IKM 和 FHO-IAP)都表现很好。这就表明在这个随机生成的大数据集上,聚类效果和两个软约束之间的冲突关系更大,更需要在这两者之间取得平衡。这一现象的原因是,随机生成的数据集中大部分工艺路线之间的相似度都很低,且相近,如图 4-5 所示。这种情况下,工艺路线选哪个当其中心,区别都不大。而对于式(2-1)中的三个部分而言,此时数量软约束占据了主导。意思是,只要通过合理的聚类就可以有效消除数量软约束的罚值,从而提高聚类效果。而其他聚类算法(即 OAP、ALC 和 FHO-OAP)在聚类过程中没有考虑数量软约束,所以软约束的罚值较高,明显降低了其综合效果。

表 4-2 100 条网状工艺路线聚类结果与方差

K	OAP	ALC	IAP	IKM	FHO-OAP	FHO-IAP
1	-90.24(0.00)	-90.24(0.00)	-90.24(0.00)	-90.24(0.00)	-90.24(0.00)	-90.24(0.00)
2	-179.99(0.00)	-188.33(0.00)	-173.64(8.61)	-168.87(0.00)	-171.73(0.00)	-171.35(0.72)
3	-269.91(0.00)	-282.32(0.00)	-255.67(0.48)	-254.43(0.00)	-259.48(0.00)	-254.66(0.25)
4	-352.89(0.00)	-358.12(0.00)	-338.60(0.74)	-338.37(0.21)	-343.48(0.00)	-337.55(0.00)
5	-425.83(0.00)	-436.98(0.00)	-420.53(0.00)	-419.57(0.98)	-425.89(0.02)	-419.44(0.44)
6	-503.90(0.00)	-516.61(0.00)	-496.86(0.00)	-498.86(2.05)	-501.82(0.31)	-495.36(0.00)
7	-560.91(0.00)	-589.71(0.00)	-558.96(0.00)	-567.21(1.98)	-560.78(0.03)	-558.70(0.00)
8	-626.98(0.00)	-668.30(0.00)	-625.20(0.00)	-637.87(1.37)	-598.36(2.15)	-598.60(3.26)
Rank	4.4375	5.6875	2.9375	2.8125	3.4375	1.6875

图 4-9 显示了六种算法在大型数据集上的聚类效果随 K 值的变化趋势。从中可以看出,随着的增加,所有算法的 *netsim* 呈现出明显下降趋势。其中理由如前所述,本节所用大数据集的大部分工艺路线间相似度都很小,彼此间差距不大。这种情况下,每个工艺路线到中心的相似度之和不会随着聚类结果出现大的差异。而图 4-9 中的彩色面积同样表示的是各个算法计算出 *netsim* 的方差(即表 4-2 括号中的数值)。由于在大数据集中,大部分工艺路线之间的相似度差异较小聚类效果的提升空间也小,故各算法所得结果的差异也相对较小,导致图中的彩色面积较小。对于半径软约束而言,因为相同的原因, K 值的变化并不会导致聚类半

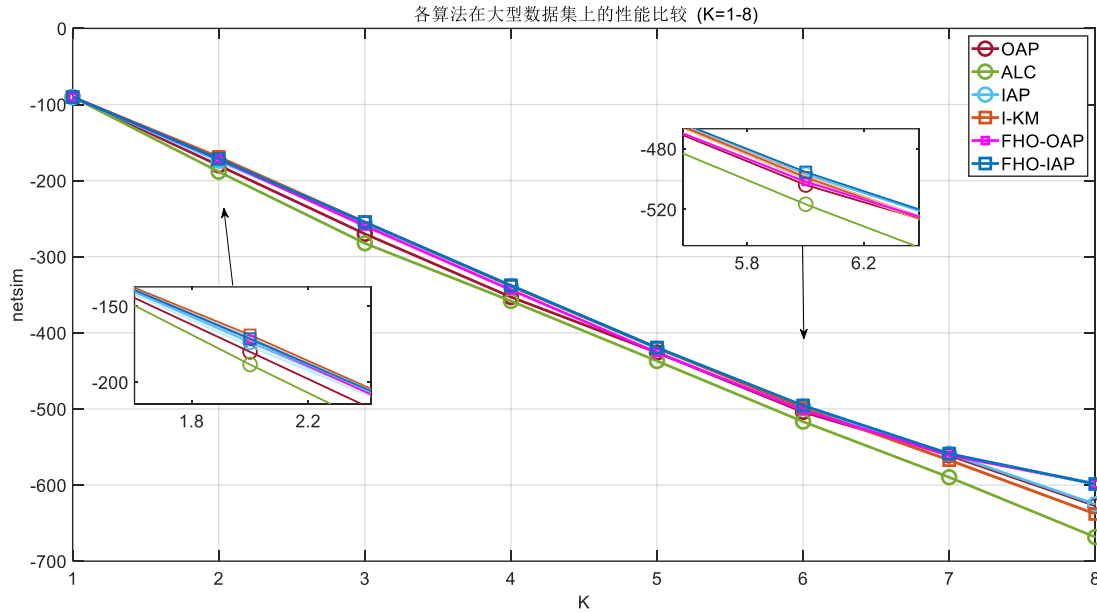


图 4-9 6 种算法对 100 条网状工艺路线聚类效果对比图

径的大幅变化，所以不同聚类结果间的差距不大。唯一受到反值影响的是数量软约束，因为其计算方式与相似度无关，只与类内个数有关。所以，随着区值的增大，数量软约束的 n 也在下降，导致了其对应的罚值在上升。如此，总的 $netsim$ 就会呈现出图 4-9 中的明显下降趋势。

仔细观察图 4-9 可以发现，FHO-IAP 算法在大型数据集上的聚类效果最佳，而 ALC 算法同样是六个算法中聚类效果最差的算法。这是因为 ALC 算法在处理大规模数据集时存在一定局限性，ALC 算法聚类中心的选择和更新机制不够优化，导致其聚类效果较差。而 FHO-IAP 算法不但在小型数据集的聚类问题处理中表现优异，在大型数据集的聚类问题中也不会因为数据集的规模大小和分布而影响聚类效果，这体现出了 FHO-IAP 算法的稳健性更好。

此外，对比图 4-7 与图 4-9 中的 ALC 聚类结果发现，图 4-7 中 ALC 的聚类效果呈现出明显的不稳定性，而图 4-9 中的 ALC 聚类效果与其他算法的聚类效果总体趋势相同。这是因为两个数据集的相似度值分布不同从而影响了数量软约束和半径软约束的罚值计算，进而在计算不同 K 值下 ALC 的 $netsim$ 时呈现出不同的变化趋势。如图 3-6 所示，小数据集因为尽可能全面地囊括了网状工艺路线的不同相似度情况，故其数据分布情况更加复杂和均衡。反映到聚类过程，就要求算法能在聚类结果和两个软约束之前取得平衡，以使式(2-1)的 $netsim$ 最大化。

由于 ALC 的聚类过程只考虑了类间平均距离，完全没有考虑两个软约束的需求故其聚类结果极易受到两个软约束罚值的影响当 $K=2, 4, 5$ 时，ALC 的聚类结果违反软约束的程度更加严重，故罚值更大，直接拉低了 ALC 的整体性能。而对于大数据集而言，由图 4-7 可知，大部分工艺路线间的相似度较小且差异不大，根据上文对图 4-9 变化趋势的解释，随着 K 值的增加，真正能引起 *netsim* 剧烈变化的只有数量软约束。在这种情况下，一方面，无论是 ALC 还是其他算法在大数据集的聚类过程中更倾向于产生类内数量均匀的聚类结果；另一方面，数量软约束的约束条件 n 随 K 值的上升而匀速下降，故对应的罚值也在上升。因此，无论是 ALC 还是其他算法在大数据集上都表现出相同的变化趋势。

综合上述两个数据集的实验结果可知，本文所提 FHO-OAP 和 FHO-IAP 相较于原始的 OAP 和 IAP 算法都有了明显的性能提升，说明 FHO 的引入确实能帮助 AP 算法自动选择最合适的参数。此外，表 4-1 和表 4-2 表明 FHO-IAP 在两个数据集中都表现出最好的聚类效果，因为其性能平均排名(Rank)都是最低。同时，FHO-IAP 所得结果的方差在六个算法中也相对较小，体现了其在不同数据分布的典型工艺路线发现问题上都具有较强的稳健性。

4.5 本章小结

本章针对典型工艺路线发现问题的需求和挑战，提出了一种结合 FHO 算法和 AP 聚类算法的创新方法，旨在解决传统聚类算法在复杂网状工艺路线分析中的局限性。基于 FHO 的全局优化能力和 AP 算法以真实数据点为聚类中心的特性，分别设计了 FHO-OAP 和 FHO-IAP 两种改进算法，通过优化参考度和阻尼系数，有效提高了聚类的质量和性能。通过仿真实验对比分析了不同算法在小型和大型网状工艺路线数据集上的表现。结果显示，FHO-OAP 和 FHO-IAP 在聚类效果、计算效率和稳定性方面均优于传统方法，尤其是 FHO-IAP 算法能够在多样化的数据场景中展现出更优的适用性和鲁棒性。

第 5 章 总结与展望

5.1 总结

在工艺规划过程中,高质量的典型工艺路线可以帮助 CAPP 检索和推送更具代表性和符合生产需求的有效信息。而在以往的研究中,甚少有学者根据网状工艺路线的特点提出相应的典型网状工艺路线发现方法。因此,本文提出了一种典型网状工艺路线的自动化发现方法:(1)是设计了一种多维工艺信息融合的新型网状工艺路线综合相似度度量,以实现更加细粒度地量化网状工艺路线间的共有工艺信息;(2)是在考虑聚类有效性的同时基于 FHO 优化 AP 来自动化发现典型网状工艺路线,以保证所得典型工艺路线的质量。

为了有效计算网状工艺路线之间的相似度,本文首先对网状工艺路线进行了深入剖析,认为其中包含了四种工艺信息:先后顺序关系、共同工序数、工序间相似度与结构复杂度,为了有效量化这四种工艺信息,本文设计了双结构矩阵和 KM 算法来同时度量工序间相似度和先后顺序关系;用了 Jaccard 相似度来衡量共同工序数,而结构复杂度则通过可选工艺路线条数来量化。最后,基于 PCA 分析了这些信息的相关性,并得到了针对网状工艺路线的综合相似度度量。

对于网状工艺路线的聚类过程,首先在考虑聚类有效性的基础上为网状工艺路线发现问题引入了数量软约束与半径软约束,以保证聚类结果的真实有效,同时,为了解决 AP 算法的参考度与阻尼系数无法自动调优问题,将 FHO 引入中。FHO 通过不断迭代计算可为 AP 算法自动选择能使聚类效果最优的参数,在保证聚类有效性的同时提高发现的典型网状工艺路线的质量。相关实验也表明,FHO 确实可以提高相应 AP 算法的聚类效果,其中 FHO-IAP 算法相较于其他算法而言具有更好的聚类效果和更强的稳健性。

5.2 展望

本文主要围绕多维工艺信息融合的典型网状工艺路线发现方法展开研究。尽管取得了一定的成果,但由于本人学识有限,该领域仍有诸多问题亟待进一步探索与解决。具体而言,以下几个方面值得深入研究:

（1）相似度度量方法的局限性。本文提出的相似度度量方法在解决重复工序的问题上仍存在不足。这一问题限制了方法在复杂工艺路线中的广泛适用性，未来可以尝试引入更多适配重复工序特性的相似度衡量机制，以增强方法的通用性和精确性。

（2）聚类倾向性问题。研究发现，聚类过程容易偏向选择结构更复杂的网状工艺路线。这是因为工艺路线的结构复杂度通常与其所蕴含的工序信息正相关。然而，过于复杂的典型工艺路线虽然能够提供更多的信息，却可能在实际应用中降低对工艺规划的辅助效果。因此，未来研究需要探索平衡复杂度与实用性的方法，或提出多层次、多尺度的工艺路线发现策略，以提升研究的实际价值。

参考文献

- [1] Xiao Y, Zheng S, Shi J, et al. Knowledge graph-based manufacturing process planning: A state-of-the-art review [J]. Journal of Manufacturing Systems, 2023, 70: 417-435.
- [2] 张超, 周光辉, 李晶晶, 等. 面向航空复杂零件智能工艺规划的孪生工艺模型构建与应用研究[J]. 机械工程学报, 2024, 60(06): 32-43.
- [3] 葛睿夫, 任志刚, 林江豪, 等. 面向注塑产品工艺缺陷的知识图谱构建方法及应用[J]. 控制理论与应用, 2024, 41(03): 577-585.
- [4] Duan J ,Duan Y .Toward a framework of extracting typical machining process routines based on knowledge representation learning [J]. Advanced Engineering Informatics, 2024, 60: 102431.
- [5] Thanos C, Meghini C, Bartalesi V, et al. An exploratory approach to data driven knowledge creation [J]. Journal of Big Data, 2023, 10(1): 29.
- [6] Zhang H, Wang W, Zhang S, et al. A novel method based on deep reinforcement learning for machining process route planning [J]. Robotics and Computer-Integrated Manufacturing, 2024, 86: 102688.
- [7] Li C. Study on typical process route mining method based on multilevel longest common subsequence information entropy and intelligent clustering model [J]. International Journal of Computer Integrated Manufacturing,2023, 36(10):1416-1430.
- [8] Liu N S, Tian T X, Zhang M Z, et al. Applying hierarchical clustering to discover the typical process route [J]. Materials Science Forum, 2006, 51(532-533): 949-952.
- [9] 张辉, 裘乐淼, 张树有, 等. 基于智能聚类分析的产品典型工艺路线提取方法[J]. 计算机集成制造系统, 2013, 19(03): 490-498.
- [10] Xin Y, Chen Y, Li W, et al. Refined simulation method for computer-aided process planning based on digital twin technology [J]. Micromachines, 2022, 13(4): 620.
- [11] 朱震宇. 离散制造系统的典型工艺知识自动化发现方法[D]. 江南大学, 2019.

- [12]Navaei J, ElMaraghy H. Minimizing backtracking distance for networked operations in smart manufacturing [J]. International Journal of Computer Integrated Manufacturing, 2021, 34(5): 515-531.
- [13]Zhou D, Dai X. Integrating granular computing and bioinformatics technology for typical process routes elicitation: a process knowledge acquisition approach [J]. Engineering Applications of Artificial Intelligence, 2015, 45: 46-56.
- [14]Liu Q, Li X, Gao L. A novel MILP model based on the topology of a network graph for process planning in an intelligent manufacturing system [J]. Engineering, 2021, 7(6): 807-817.
- [15]Newman S T, Zhu Z, Dhokia V, et al. Process planning for additive and subtractive manufacturing technologies [J]. CIRP annals, 2015, 64(1): 467-470.
- [16]Tan W, Khoshnevis B. Integration of process planning and scheduling—a review [J]. Journal of Intelligent Manufacturing, 2000, 11: 51-63.
- [17]Kumar S P L. State of the art-intense review on artificial intelligence systems application in process planning and manufacturing[J]. Engineering Applications of Artificial Intelligence, 2017, 65: 294-329.
- [18]Yu Z, Fan D, Sha L, et al. Process planning of repairing and grinding for cylindrical parts based on surface defect detection [J]. Measurement Science and Technology, 2024, 36(1): 015434.
- [19]Zhang L, Wu H, Chen Y, et al. A knowledge-guided process planning approach with reinforcement learning [J]. Journal of Engineering Design, 2024: 1-24.
- [20]Wang J, Zheng H, Zhao S, et al. Energy-aware remanufacturing process planning and scheduling problem using reinforcement learning-based particle swarm optimization algorithm [J]. Journal of Cleaner Production, 2024, 476: 143771.
- [21]Cheng K, Zhao G, Wang W, et al. A STEP-NC complaint and feature-based solution for intelligent process planning [J]. The International Journal of Advanced Manufacturing Technology, 2024, 133(11): 6023-6039.

- [22]Guo K, Liu R, Duan G, et al. A Deep reinforcement learning method with multiple starting nodes for dynamic process planning decision making [J]. Computers & Industrial Engineering, 2024, 194: 110359.
- [23]Liang J, Zhang S, Zhang Y, et al. A knowledge graph-based approach to modeling & representation for machining process design intent [J]. Advanced Engineering Informatics, 2024, 62: 102645.
- [24]Choobineh F. A framework for the design of cellular manufacturing systems [J]. The International Journal of Production Research, 1988, 26(7): 1161-1172.
- [25]Zhou D, Dai X. A method for discovering typical process sequence using granular computing and similarity algorithm based on part features [J]. The International Journal of Advanced Manufacturing Technology, 2015, 78: 1781-1793.
- [26]Liu S, Zhang Z, Tian X. A typical process route discovery method based on clustering analysis [J]. The International Journal of Advanced Manufacturing Technology, 2007, 35: 186-194.
- [27]Xu B, Wang Y, Ji Z. A novel operation sequence similarity-based approach for typical process route knowledge discovery [J]. IEEE Access, 2021, 9: 126801-126821.
- [28]Navaei J, ElMaraghy H. Grouping part/product variants based on networked operations sequence [J]. Journal of Manufacturing Systems, 2016, 38: 63-76.
- [29]Noshari R M, Azgomi H, Asghari A. Efficient clustering in data mining applications based on harmony search and k-medoids[J]. Soft Computing, 2024, 28(23): 1-24.
- [30]Hamasuna Y, Nakano S, Ozaki R, et al. Cluster validity measures based agglomerative hierarchical clustering for network data [J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2019, 23(3): 577-583.
- [31]Huang K. The use of a newly developed algorithm of divisive hierarchical clustering for remote sensing image analysis [J]. International Journal of Remote Sensing, 2002, 23(16): 3149-3168.

- [32]Wu L, Suzuki S. Cell formation design with improved similarity coefficient method and decomposed mathematical model [J]. The International Journal of Advanced Manufacturing Technology, 2015, 79: 1335-1352.
- [33]王琳,张永健, 钟诗胜. 基于近邻传播聚类的卫星典型构件典型工艺过程发现[J]. 计算机集成制造系统, 2015, 21(06): 1469-1475.
- [34]Gupta T, Seifoddini H I. Production data based similarity coefficient for machine-component grouping decisions in the design of a cellular manufacturing system [J]. The International Journal of Production Research, 1990, 28(7): 1247-1269.
- [35]徐彬梓. 基于知识的离散制造系统能效优化方法研究[D]. 江南大学, 2019.
- [36]Guo Y, Li W, Xiao L, et al. A prediction-based iterative Kuhn-Munkres approach for service vehicle reallocation in ride-hailing [J]. International Journal of Production Research, 2024, 62(10): 3690-3715.
- [37]Jaccard P. The distribution of the flora in the alpine zone. 1 [J]. New phytologist, 1912, 11(2): 37-50.
- [38]Azizi M, Talatahari S, Gandomi A H. Fire Hawk Optimizer: A novel metaheuristic algorithm [J]. Artificial Intelligence Review, 2023, 56(1): 287-363.
- [39]Li C. A novel framework for discovery and reuse of typical process route driven by symbolic entropy and intelligent optimisation algorithm[J]. Plos one, 2022, 17(9): e0274532.
- [40]花豹, 周彬, 顾星海, 等. 基于图神经网络的工艺表格语义相似性度量[J]. 计算机集成制造系统, 2022, 28(12): 3805-3821.
- [41]Wang L, Zhang Y, Zhong S. Typical process discovery based on affinity propagation [J]. Journal of Advanced Mechanical Design, Systems, and Manufacturing, 2016, 10(1): 1-13.
- [42]Askin R G, Zhou M. Formation of independent flow-line cells based on operation requirements and machine capabilities [J]. IIE transactions, 1998, 30(4): 319-329.
- [43]Ho Y C, Lee C E C, Moodie C L. Two sequence-pattern, matching-based, flow analysis methods for multi-flowlines layout design [J]. The International Journal of Production Research, 1993, 31(7): 1557-1578.

- [44]罗年猛, 李雄. 基于典型工艺的相似工艺路线检索方法[J]. 机械工程与自动化, 2014, (05): 101-103.
- [45]李雄飞, 李军. 数据挖掘与知识发现[M]. 北京: 高等教育出版社, 2003.
- [46]Kuhn H W. The Hungarian method for the assignment problem [J]. Naval research logistics quarterly, 1955, 2(1 - 2): 83-97.
- [47]Givoni I E, Frey B J. A binary variable model for affinity propagation [J]. Neural Computation, 2009, 21(6): 1589-1600.
- [48]Martins A S C, de Araujo L R, Penido D R R. K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow [J]. International Journal of Electrical Power & Energy Systems, 2024, 157: 109861.
- [49]Jia W, Zhao K, Zhao S. The hierarchical clustering of human mobility behaviors [J]. IEEE Transactions on Computational Social Systems, 2024, 11(2): 1876-1887.
- [50]赵嘉, 陈磊, 吴润秀, 等. K 近邻和加权相似性的密度峰值聚类算法[J]. 控制理论与应用, 2022, 39(12): 2349-2357.
- [51]Ding L, Li C, Jin D, et al. Survey of spectral clustering based on graph theory [J]. Pattern Recognition, 2024: 110366.
- [52]Frey B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007, 315(5814): 972-976.
- [53]Wang G, Huang S, Shang X, et al. Formation of part family for reconfigurable manufacturing systems considering bypassing moves and idle machines [J]. Journal of Manufacturing Systems, 2016, 41: 120-129.
- [54]Yu J, Cheng Q. The upper bound of the optimal number of clusters in fuzzy clustering [J]. Science in China Series: Information Sciences, 2001, 44: 119-125.
- [55]Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance [J]. Journal of the American Statistical Association, 1937, 32(200): 675-701.

攻读硕士期间取得的学术成果

发表的论文

- [1] Wang Mengjiao, Xu Binzi*, Huang Dengchao, Zhang Yao, Wang Chun, Yang Long. A Novel Comprehensive Similarity Measure for Networked Operation Sequences [C]. 2023 China Automation Congress (CAC), Chongqin, China, pp.1766 – 1771, 2023.11.17-2023.11.19. (EI 会议, 已录用, 对应本文第三章)
- [2] 王梦娇, 徐彬梓*, 黄登朝, 王春, 王艳. 多维工艺信息融合的典型网状工艺路线自动化发现 [J]. 控制理论与应用. (EI 收录, 已录用, 对应本文第四章)

参与的项目

- [1] 安徽省中青年教师培养行动(参与, 项目编号: JNFX2023017): 面向多源能耗的复杂动态柔性制造系统调度规则自动化生成。
- [2] 芜湖市科技项目应用基础研究项目(参与, 项目编号: 2023jc05): 面向动态柔性制造系统的多源能耗集成式协同优化。
- [3] 高等学校自然科学研究项目重点项目(参与, 项目编号: KJ2020A0345): 基于多粒度协同的离散制造系统能耗智能优化方法研究。

致谢

快乐的时光总是短暂的。硕士的三年是我求学之路最快乐的三年，也是我收获颇多的三年。在这段充满挑战和成长时光里，我不仅收获了知识和经验，更是得到了许多人无私的帮助与支持。回首这一路走来的点点滴滴，心中满是感激与温暖。在这篇硕士论文即将完成之际，我衷心想向在此过程中给予我支持和帮助的所有人表达我的感激之情。

首先，非常感谢我的导师徐彬梓老师。您严谨的学术态度、深邃的学术视野和细致入微的指导，使我在研究过程中受益匪浅。在论文写作的过程中，您不仅为我提供了宝贵的研究思路 and 方向，还在我遇到困惑和挑战时，耐心指导并鼓励我。没有您的支持与引领，我无法顺利完成这篇论文。再次表达对徐老师由衷的感恩之情。

其次，我还要感谢我的同学和朋友们。感谢费宝林师兄在实验过程中对我的指导与帮助；感谢师弟师妹曹新宇、许凯以及我的室友们皇甫紫薇、陈欢、何琴一直以来在生活中与学业上对我的帮助与支持。

最后，衷心感谢我的父母和爷爷奶奶。是你们无私的爱与支持，让我有了追求学术梦想的勇气与力量。无论在我遇到困难还是疲惫时，你们始终是我坚强的后盾。你们的理解与包容，是我坚持到最后的动力。

这篇论文的完成，是我硕士生涯的一个阶段性成果，也是我学术成长的见证。未来的路还很长，我将继续带着您们的期许和鼓励，勇敢地走下去，不断追求卓越，不负时代、不负自己。

再次感谢所有曾经给予我帮助的人！

附录

附录 A 文献[28]设计相似度度量计算结果

附录 B 本文设计相似度度量计算结果

附录 C 文献[11]设计相似度度量计算结果

附录 A 文献[28]设计相似度度量计算结果 ($\times 10^{-2}$)

	O	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
O	M	M	M	60	83	95	60	70	71	31	43	20	20	20	20	20	10	10	6	56	36	36	54	36	36	50	24	21	20	20	20	44	23	31	10	10	10	23	20	10
01	M	M	M	60	83	95	60	70	71	31	43	20	20	20	20	20	10	10	6	56	36	36	54	36	36	50	24	21	20	20	20	44	23	31	10	10	10	23	20	10
02	60	60	M	64	60	95	54	39	55	51	20	20	20	20	20	20	10	10	6	36	56	46	36	52	44	24	50	32	20	20	20	31	23	41	10	10	10	23	20	10
03	83	83	64	M	80	64	85	71	31	73	20	20	20	20	20	20	10	10	6	56	36	36	54	36	36	50	24	47	20	20	20	44	23	31	10	10	10	23	20	10
04	95	95	60	80	M	60	68	68	31	43	20	20	20	20	20	20	10	10	6	54	36	36	54	36	36	47	24	21	20	20	20	44	23	31	10	10	10	23	20	10
05	60	60	95	64	60	M	58	39	53	50	20	20	20	20	20	20	10	10	6	36	52	44	36	60	43	24	47	31	20	20	20	31	23	40	10	10	10	23	20	10
06	70	70	54	85	68	58	M	78	40	60	10	10	10	10	10	10	20	20	15	42	26	26	41	34	26	56	33	53	10	10	10	31	14	21	20	20	19	13	10	20
07	71	71	39	71	68	39	78	M	48	22	10	10	10	10	10	10	20	20	15	50	24	24	47	24	24	64	37	20	10	10	10	38	10	18	20	20	15	14	10	20
08	31	31	55	31	31	53	40	48	M	16	6	6	6	6	6	6	15	15	20	18	31	31	18	29	29	30	43	24	6	6	6	13	6	29	15	15	20	9	6	15
09	43	43	51	73	43	50	60	22	16	M	20	20	20	20	20	20	10	10	6	24	24	24	24	24	24	10	10	62	20	20	20	24	24	24	10	10	11	20	20	10
10	20	20	20	20	20	20	10	10	6	20	M	60	83	95	60	80	71	39	29	20	20	20	20	20	20	10	10	6	56	36	36	23	31	23	50	24	10	20	20	10
11	20	20	20	20	20	20	10	10	6	20	60	M	64	60	95	64	39	71	37	20	20	20	20	20	20	10	10	6	36	56	46	23	54	23	24	50	10	20	20	10
12	20	20	20	20	20	20	10	10	6	20	83	64	M	80	64	95	71	39	59	20	20	20	20	20	20	10	10	6	56	36	36	23	31	23	50	24	10	20	20	10
13	20	20	20	20	20	20	10	10	6	20	95	60	80	M	60	78	68	39	29	20	20	20	20	20	20	10	10	6	54	36	36	23	31	23	47	24	10	20	20	10
14	20	20	20	20	20	20	10	10	6	20	60	95	64	60	M	68	39	68	36	20	20	20	20	20	20	10	10	6	36	52	44	23	54	23	24	47	10	20	20	10
15	20	20	20	20	20	20	10	10	6	20	80	64	95	78	68	M	68	39	55	20	20	20	20	20	20	10	10	6	52	36	36	23	31	23	47	24	10	20	20	10
16	10	10	10	10	10	10	20	20	15	10	71	39	71	68	39	68	M	60	26	10	10	10	10	10	10	20	15	50	24	24	14	24	14	64	37	15	10	10	20	
17	10	10	10	10	10	10	20	20	15	10	39	71	39	39	68	39	60	M	26	10	10	10	10	10	10	20	20	15	24	50	37	14	47	14	37	64	15	10	10	20
18	6	6	6	6	6	6	15	15	20	6	29	37	59	29	36	55	26	26	M	6	6	6	6	6	6	15	15	20	10	10	10	6	6	6	15	15	26	6	6	15
19	56	56	36	56	54	36	42	50	18	24	20	20	20	20	20	20	10	10	6	M	60	66	95	60	65	71	39	15	44	31	31	81	36	50	18	18	10	45	42	18
20	36	36	56	36	36	52	26	24	31	24	20	20	20	20	20	20	10	10	6	60	M	69	60	95	68	39	71	15	31	44	31	50	63	59	18	18	10	35	31	18
21	36	36	46	36	36	44	26	24	31	24	20	20	20	20	20	20	10	10	6	66	69	M	65	68	95	47	46	15	31	31	44	56	42	79	18	18	10	35	31	18
22	54	54	36	54	54	36	41	47	18	24	20	20	20	20	20	20	10	10	6	95	60	65	M	60	65	68	39	15	42	31	31	81	36	50	18	18	10	43	42	18
23	36	36	52	36	36	60	34	24	29	24	20	20	20	20	20	20	10	10	6	60	95	68	60	M	68	39	68	15	31	44	31	50	63	58	18	18	10	35	31	18
24	36	36	44	36	36	43	26	24	29	24	20	20	20	20	20	20	10	10	6	65	68	95	65	68	M	46	45	15	31	31	44	56	42	84	18	18	10	35	31	18
25	50	50	24	50	47	24	56	64	30	10	10	10	10	10	10	10	20	20	15	71	39	47	68	39	46	M	60	20	18	18	18	50	18	30	30	30	20	22	18	30
26	24	24	50	24	24	47	33	37	43	10	10	10	10	10	10	10	20	20	15	39	71	46	39	68	45	60	M	20	18	18	18	30	18	38	30	30	20	22	18	30
27	21	21	32	47	21	31	53	20	24	62	6	6	6	6	6	6	15	15	20	15	15	15	15	15	15	20	20	M	10	10	10	15	15	15	20	20	60	9	10	20
28	20	20	20	20	20	20	10	10	6	20	56	36	56	54	36	52	50	24	10	44	31	31	42	31	31	18	18	10	M	60	66	52	50	46	71	39	15	43	42	18
29	20	20	20	20	20	20	10	10	6	20	36	56	36	36	52	36	24	50	10	31	44	31	31	44	31	18	18	10	60	M	69	36	81	36	39	71	15	30	31	18
30	20	20	20	20	20	20	10	10	6	20	36	46	36	36	44	36	24	37	10	31	31	44	31	31	44	18	18	10	66	69	M	36	55	56	47	46	15	30	31	18
31	44	44	31	44	44	31	31	38	13	24	23	23	23	23	23	23	14	14	6	81	50	56	81	50	56	50	30	15	52	36	36	M	42	42	35	24	10	43	42	18
32	23	23	23	23	23	23	14	10	6	24	31	54	31	31	54	31	24	47	6	36	63	42	36	63	42	18	18	15	50	81	55	42	M	48	39	68	10	30	31	18
33	31	31	41	31	31	40	21	18	29	24	23	23	23	23	23	23	14	14	6	50	59	79	50	58	84	30	38	15	46	36	56	42	48	M	37	24	10	30	31	18
34	10	10	10	10	10	10	20	20	15	10	50	24	50	47	24	47	64	37	15	18	18	18	18	18	18	30	30	20	71	39	47	35	39	37	M	60	20	18	18	30
35	10	10	10	10	10	10	20	20	15	10	24	50	24	24	47	24	37	64	15	18	18	18	18	18	18	30	30	20	39	71	46	24	68	24	60	M	20	18	18	30
36	10	10	10	10	10	10	19	15	20	11	10	10	10	10	10	10	15	15	26	10	10	10	10	10	10	20	20	60	15	15	15	10	10	10	20	20	M	9	10	20
37	23	23	23	23	23	23	13	14	9	20	20	20	20	20	20	20	10	10	6	45	35																			

*O 是指原始路线，M 是指两条工艺路线之间最大相似度 1，01-39 是指图 2-2 中的相应相似度情况。

附录 B 本文设计相似度度量计算结果 ($\times 10^{-2}$)

	O	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	
O	M	M	M	69	88	99	76	90	73	58	48	28	13	22	31	19	26	25	17	17	49	42	42	52	45	45	48	40	38	15	13	12	44	23	35	20	17	24	17	6	10
01	M	M	M	69	88	M	76	90	73	58	48	28	13	22	31	19	25	25	17	17	49	42	42	52	45	45	48	40	38	15	13	12	44	23	35	20	17	24	17	6	10
02	69	69	M	77	76	M	79	58	73	51	13	28	15	19	31	20	17	25	17	42	49	42	45	52	48	40	48	42	13	15	13	37	26	40	17	20	27	17	6	10	
03	88	88	77	M	90	79	M	73	58	59	22	15	28	25	20	31	25	17	22	49	42	42	52	45	45	48	40	46	15	13	12	44	23	35	20	17	30	17	6	10	
04	M	M	M	76	90	M	73	87	73	59	49	31	19	25	28	13	20	26	20	20	52	45	45	49	39	42	48	42	40	20	18	18	40	17	32	22	20	26	19	0	13
05	76	76	M	79	73	M	80	59	73	53	19	31	20	13	28	17	20	26	20	45	52	45	39	52	42	42	48	44	18	20	18	31	24	36	20	22	29	19	0	13	
06	90	90	79	M	87	80	M	73	59	59	25	20	31	20	17	28	26	20	24	52	45	45	45	45	42	48	42	47	20	18	17	37	21	32	22	20	31	19	0	13	
07	73	73	58	73	73	59	73	M	73	27	25	17	25	26	20	26	27	13	9	48	40	40	48	42	42	60	47	13	20	17	16	40	20	31	18	13	4	8	13	0	
08	58	58	73	58	59	73	59	73	M	13	17	25	17	20	26	20	13	27	0	40	48	40	42	48	45	47	60	6	17	20	17	33	22	37	13	18	0	8	13	0	
09	48	48	51	59	49	53	59	27	13	M	17	17	22	20	20	24	9	0	30	22	22	22	24	24	24	0	0	73	15	15	15	24	24	24	0	0	42	0	13	0	
10	28	28	13	22	31	19	25	25	17	17	M	69	88	M	76	90	73	58	48	15	13	12	20	18	18	20	17	16	49	42	42	26	37	26	48	40	22	11	6	10	
11	13	13	28	15	19	31	20	17	25	17	69	M	77	76	78	79	58	73	51	13	15	13	18	20	19	17	20	17	42	49	42	23	44	27	40	48	24	11	6	10	
12	22	22	15	28	25	20	31	25	17	22	88	77	M	90	79	M	73	58	59	15	13	12	20	18	18	20	17	20	49	42	42	26	37	26	48	40	27	11	6	10	
13	31	31	19	25	28	13	20	26	20	20	M	76	90	M	73	87	73	59	49	20	18	18	16	11	12	22	20	18	52	45	45	23	31	20	48	42	25	14	0	13	
14	19	19	31	20	13	28	17	20	26	20	76	78	79	73	M	80	59	73	53	18	20	18	11	17	13	20	22	20	45	52	45	17	43	20	42	48	26	14	0	13	
15	25	25	20	31	20	17	28	26	20	24	90	79	M	87	80	M	73	59	59	20	18	17	13	15	12	22	20	22	52	45	45	19	36	20	48	42	29	14	0	13	
16	25	25	17	25	26	20	26	27	13	9	73	58	73	73	59	73	M	73	27	20	17	16	22	20	20	18	13	4	48	40	40	28	42	28	60	47	4	1	13	0	
17	17	17	25	17	20	26	20	13	27	0	58	73	58	59	73	59	73	M	13	17	20	17	20	22	21	13	18	0	40	48	40	26	48	29	47	60	0	1	13	0	
18	17	17	17	22	20	20	24	9	0	30	48	51	59	49	53	59	27	13	M	15	15	15	17	17	17	0	0	24	22	22	22	17	17	17	0	0	36	0	13	0	
19	49	49	42	49	52	45	52	48	40	22	15	13	15	20	18	20	20	17	15	M	69	77	78	76	79	73	58	31	51	28	36	86	42	66	37	27	23	49	39	29	
20	42	42	49	42	45	52	45	40	48	22	13	15	13	18	20	18	17	20	15	69	M	77	76	M	83	58	73	28	28	51	33	62	61	69	27	37	19	42	30	24	
21	42	42	42	42	45	45	45	40	40	22	12	13	12	18	18	17	16	17	15	77	77	M	79	79	M	62	58	31	36	33	51	67	45	84	32	27	23	47	35	29	
22	52	52	45	52	49	39	45	48	42	24	20	18	20	16	11	13	22	20	17	78	76	79	M	73	76	73	59	33	53	35	40	85	39	63	38	29	25	50	36	31	
23	45	45	52	45	39	52	45	42	48	24	18	20	18	11	17	15	20	22	17	76	M	79	73	M	76	59	73	30	35	53	37	59	59	65	29	38	22	43	24	27	
24	45	45	48	45	42	42	42	42	45	24	18	19	18	12	13	12	20	21	17	79	83	M	76	76	M	63	63	33	39	38	53	64	42	84	33	30	25	48	28	31	
25	48	48	40	48	48	42	48	60	47	0	20	17	20	22	20	22	18	13	0	73	58	62	73	59	63	M	73	13	37	27	32	61	29	50	43	25	13	22	29	25	
26	40	40	48	40	42	48	42	47	60	0	17	20	17	20	22	20	13	18	0	58	73	58	59	73	63	73	M	6	27	37	27	47	38	50	25	43	6	20	25	18	
27	38	38	42	46	40	44	47	13	6	73	16	17	20	18	20	22	4	0	24	31	28	31	33	30	33	13	6	M	23	19	23	33	30	33	13	6	60	8	21	13	
28	15	15	13	15	20	18	20	20	17	15	49	42	49	52	45	52	48	40	22	51	28	36	53	35	39	37	27	23	M	69	77	61	63	50	73	58	31	44	39	29	
29	13	13	15	13	18	20	18	17	20	15	42	49	42	45	52	45	40	48	22	28	51	33	35	53	38	27	37	19	69	M	77	43	86	48	58	73	28	34	30	24	
30	12	12	13	12	18	18	17	16	17	15	42	42	42	45	45	45	40	40	22	36	33	51	40	37	53	32	27	23	77	77	M	48	66	63	62	58	31	39	35	29	
31	44	44	37	44	40	31	37	40	33	24	26	23	26	23	17	19	28	26	17	86	62	67	85	59	64	61	47	33	61	43	48	M	48	53	46	37	25	50	36	31	
32	23	23	26	23	17	24	21	20	22	24	37	44	37	31	43	36	42	48	17	42	61	45	39	59	42	29	38	30	63	86	66	48	M	53	59	73	22	36	24	27	
33	35	35	40	35	32	36	32	31	37	24	26	27	26	20	20	20	28	29	17	66	69	84	63	65	84	50	50	33	50	48	63	53	53	M	44	41	25	40	28	31	
34	20	20	17	20	22	20	22	18	13	0	48	40	48	48	42	48	60	47	0	37	27	32	38	29	33	43	25	13	73	58	62	46	59	44	M	73	13	16	29	25	
35	17	17	20	17	20	22	20	13	18	0	40	48	40	42	48	42	47	60	0	27	37	27	29	38	30	25	43	6	58	73	58	37	73	41	73	M	6	12	25	18	
36	24	24	27	30	26	29	31	4	0	42	22	24	27	25	26	29	4	0	36	23	19	23	25	22	25	13	6	60	31	28	31	25	22	25	13	6	M	8	21	13	
37	17	17	17	17	19	19	8	8	0	11	11	11	14	14	14	1	1	0	49	42	47	50	43	48	22	20	8	44	34	39	5										

*O 是指原始路线，M 是指两条工艺路线之间最大相似度 1，01-39 是指图 2-2 中的相应相似度情况。

附录 C 文献[11]设计相似度度量计算结果 ($\times 10^{-2}$)

	O	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
O	M	M	M	60	83	95	60	70	71	34	43	20	20	20	20	20	10	10	6	48	36	36	47	36	36	40	24	21	20	20	20	36	23	31	10	10	10	23	20	10
01	M	M	M	60	83	95	60	70	71	34	43	20	20	20	20	20	10	10	6	48	36	36	47	36	36	40	24	21	20	20	20	36	23	31	10	10	10	23	20	10
02	60	60	M	62	60	95	52	39	66	46	20	20	20	20	20	20	10	10	6	36	48	40	36	46	39	24	40	25	20	20	20	31	23	35	10	10	10	23	20	10
03	83	83	62	M	80	62	85	71	34	67	20	20	20	20	20	20	10	10	6	48	36	36	47	36	36	40	24	37	20	20	20	36	23	31	10	10	10	23	20	10
04	95	95	60	80	M	60	68	68	34	43	20	20	20	20	20	20	10	10	6	47	36	36	47	36	36	38	24	21	20	20	20	36	23	31	10	10	10	23	20	10
05	60	60	95	62	60	M	54	39	63	46	20	20	20	20	20	20	10	10	6	36	46	39	36	50	39	24	38	25	20	20	20	31	23	34	10	10	10	23	20	10
06	70	70	52	85	68	54	M	78	43	49	10	10	10	10	10	10	20	20	15	36	26	26	35	30	26	47	33	39	10	10	10	25	14	21	20	20	19	13	10	20
07	71	71	39	71	68	39	78	M	55	22	10	10	10	10	10	10	20	20	15	40	24	24	38	24	24	53	37	20	10	10	10	26	10	18	20	20	15	14	10	20
08	34	34	66	34	34	63	43	55	M	17	6	6	6	6	6	6	15	15	20	19	35	24	19	33	24	32	48	25	6	6	6	14	6	20	15	15	20	9	6	15
09	43	43	46	67	43	46	49	22	17	M	20	20	20	20	20	20	10	10	6	24	24	24	24	24	24	10	10	54	20	20	20	24	24	24	10	10	11	20	20	10
10	20	20	20	20	20	20	10	10	6	20	M	60	83	95	60	80	71	39	29	20	20	20	20	20	10	10	6	48	36	36	23	31	23	40	24	10	20	20	10	
11	20	20	20	20	20	20	10	10	6	20	60	M	62	60	95	62	39	71	32	20	20	20	20	20	20	10	10	6	36	48	40	23	45	23	24	40	10	20	20	10
12	20	20	20	20	20	20	10	10	6	20	83	62	M	80	62	95	71	39	53	20	20	20	20	20	20	10	10	6	48	36	36	23	31	23	40	24	10	20	20	10
13	20	20	20	20	20	20	10	10	6	20	95	60	80	M	60	78	68	39	29	20	20	20	20	20	20	10	10	6	47	36	36	23	31	23	38	24	10	20	20	10
14	20	20	20	20	20	20	10	10	6	20	60	95	62	60	M	63	39	68	31	20	20	20	20	20	20	10	10	6	36	46	39	23	45	23	24	38	10	20	20	10
15	20	20	20	20	20	20	10	10	6	20	80	62	95	78	63	M	68	39	45	20	20	20	20	20	20	10	10	6	46	36	36	23	31	23	38	24	10	20	20	10
16	10	10	10	10	10	10	20	20	15	10	71	39	71	68	39	68	M	60	26	10	10	10	10	10	10	20	20	15	40	24	24	14	24	14	53	37	15	10	10	20
17	10	10	10	10	10	10	20	20	15	10	39	71	39	39	68	39	60	M	26	10	10	10	10	10	10	20	20	15	24	40	29	14	38	14	37	53	15	10	10	20
18	6	6	6	6	6	6	15	15	20	6	29	32	53	29	31	45	26	26	M	6	6	6	6	6	6	15	15	20	10	10	10	6	6	6	15	15	26	6	6	15
19	48	48	36	48	47	36	36	40	19	24	20	20	20	20	20	20	10	10	6	M	60	63	95	60	63	71	39	15	39	31	31	75	36	50	18	18	10	41	38	18
20	36	36	48	36	36	46	26	24	35	24	20	20	20	20	20	20	10	10	6	60	M	64	60	95	63	39	71	15	31	39	31	50	57	54	18	18	10	35	31	18
21	36	36	40	36	36	39	26	24	24	24	20	20	20	20	20	20	10	10	6	63	64	M	63	63	95	44	42	15	31	31	39	54	38	73	18	18	10	35	31	18
22	47	47	36	47	47	36	35	38	19	24	20	20	20	20	20	20	10	10	6	95	60	63	M	60	63	68	39	15	38	31	31	75	36	50	18	18	10	40	38	18
23	36	36	46	36	36	50	30	24	33	24	20	20	20	20	20	20	10	10	6	60	95	63	60	M	63	39	68	15	31	36	31	50	52	53	18	18	10	35	31	18
24	36	36	39	36	36	39	26	24	24	24	20	20	20	20	20	20	10	10	6	63	63	95	63	63	M	43	42	15	31	31	36	53	38	84	18	18	10	35	31	18
25	40	40	24	40	38	24	47	53	32	10	10	10	10	10	10	10	20	20	15	71	39	44	68	39	43	M	60	20	18	18	18	46	18	30	30	30	20	22	18	30
26	24	24	40	24	24	38	33	37	48	10	10	10	10	10	10	10	20	20	15	39	71	42	39	68	42	60	M	20	18	18	18	30	18	33	30	30	20	22	18	30
27	21	21	25	37	21	25	39	20	25	54	6	6	6	6	6	6	15	15	20	15	15	15	15	15	15	20	20	M	10	10	10	15	15	15	20	20	47	9	10	20
28	20	20	20	20	20	20	10	10	6	20	48	36	48	47	36	46	40	24	10	39	31	31	38	31	31	18	18	10	M	60	63	46	50	42	71	39	15	38	38	18
29	20	20	20	20	20	20	10	10	6	20	36	48	36	36	46	36	24	40	10	31	39	31	31	36	31	18	18	10	60	M	64	36	81	36	39	71	15	30	31	18
30	20	20	20	20	20	20	10	10	6	20	36	40	36	36	39	36	24	29	10	31	31	39	31	31	36	18	18	10	63	64	M	36	52	48	44	42	15	30	31	18
31	36	36	31	36	36	31	25	26	14	24	23	23	23	23	23	23	14	14	6	75	50	54	75	50	53	46	30	15	46	36	36	M	42	42	31	24	10	40	38	18
32	23	23	23	23	23	23	14	10	6	24	31	45	31	31	45	31	24	38	6	36	57	38	36	52	38	18	18	15	50	81	52	42	M	44	39	68	10	30	31	18
33	31	31	35	31	31	34	21	18	20	24	23	23	23	23	23	23	14	14	6	50	54	73	50	53	84	30	33	15	42	36	48	42	44	M	32	24	10	30	31	18
34	10	10	10	10	10	10	20	20	15	10	40	24	40	38	24	38	53	37	15	18	18	18	18	18	18	30	30	20	71	39	44	31	39	32	M	60	20	18	18	30
35	10	10	10	10	10	10	20	20	15	10	24	40	24	24	38	24	37	53	15	18	18	18	18	18	18	30	30	20	39	71	42	24	68	24	60	M	20	18	18	30
36	10	10	10	10	10	10	19	15	20	11	10	10	10	10	10	10	15	15	26	10	10	10	10	10	10	20	20	47	15	15	15	10	10	10	20	20	M	9	10	20
37	23	23	23	23	23	23	13	14	9	20	20	20	20	20	20	20	10	10	6	41	35	35	40	35	35	22	22	9												

*O 是指原始路线，M 是指两条工艺路线之间最大相似度 1，01-39 是指图 2-2 中的相应相似度情况。