成品油管道运行多参数时空模式提取与可视化

纪连恩[1,2], 邢智博[1,2], 吴昆[1,2], 赵伟[3]

[1] (中国石油大学(北京)石油数据挖掘北京市重点实验室 北京 102249)
[2] (中国石油大学(北京)计算机科学与技术系 北京 102249)
[3] (中国石油大学(北京)城市油气输配技术北京市重点实验室 北京 102249)
(jilianen@cup.edu.cn)

摘 要: 成品油管道运行过程不仅具有典型的时空特点, 且其运行模式需要由多个监测参数综合表征. 针对现有的时空模式分析方法难以揭示多参数的综合时空特征的问题, 提出一种基于多参数融合的张量分解方法用于成品油管道运行多参数时空模式的提取. 首先根据不同分析角度, 通过对管道运行的多维监测参数进行信息量及相关性分析实现分组融合; 然后将融合后的时空数据建模为张量, 使用张量分解及聚类的方法获取数据集的多维时空模式; 最后对不同模式下原始多参数变化趋势的对比分析, 进一步发现运行模式的时空规律. 基于所提方法设计了一套可视化系统 MPVis, 以支持用户从不同分析角度对多参数表征的综合时空模式进行提取及可视化. 通过真实成品油管道数据的案例结果表明, 该方法为后续成品油管道数据分析提供了一种新思路.

关键词: 时空模式可视化; 参数融合; 成品油管道; 张量分解
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00613

Extraction and Visualization of Multi-Parameter Spatio-Temporal Pattern of Multiproduct Pipeline Operation

Ji Lianen[1,2], Xing Zhibo[1,2], Wu Kun[1,2], and Zhao Wei[3]

[1] (*Beijing Key Laboratory of Petroleum Data Mining, China University of Petroleum, Beijing* 102249)
[2] (*Department of Computer Science and Technology, China University of Petroleum, Beijing* 102249)
[3] (*Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum, Beijing* 102249)

Abstract: The operation process of a multiproduct oil pipeline not only has typical spatio-temporal characteristics, but also its operation mode needs to be comprehensively characterized by multiple monitoring parameters. However, the existing spatio-temporal pattern analysis methods make it difficult to reveal the comprehensive spatio-temporal characteristics of multiple parameters. Therefore, a tensor decomposition method based on multi-parameter fusion is proposed to extract the multi-parameter spatio-temporal pattern of multiproduct pipeline operation. Firstly, this method realizes group fusion by analyzing the amount of information and correlations of multi-dimensional monitoring parameters of pipeline operation from different analytical perspectives, and then models the fused spatio-temporal data as tensors and uses tensor decomposition and clustering methods to obtain the multi-dimensional spatio-temporal patterns of the data set. Finally, through comparing the changing trend of the original multiple parameters under different patterns, the spatial and temporal law of the operation pattern is further found. Based on this method, a visualization system MPVis is designed to extract and visualize the com-

收稿日期: 2023-09-16; 修回日期: 2024-04-24. **基金项目**: 国家自然科学基金(60873093, 61972353); 中国石油天然气集团有限公司-中国石油大学(北京)战略合作科技专项(ZLZX2020-05). **纪连恩**(1972—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为可视化、人机交互、计算机图形学; **邢智博**(1998—), 男, 硕士研究生, 主要研究方向为可视化; **吴昆**(1999—), 男, 硕士研究生, 主要研究方向为可视化; **赵伟**(1997—), 男, 博士研究生, 主要研究方向为成品油管道运输策略.

prehensive spatial and temporal pattern of multi-parameter representation from different analytical perspectives. The results from case studies on real-world data show that the method provides a new thinking way for the later in-depth analysis of multiproduct oil pipeline data.

Key words: spatio-temporal pattern visualization; parameter fusion; multiproduct pipeline; tensor decomposition

成品油管道主要用于城市用油供给和工业用油供给, 管道长时间运行产生的历史数据属于典型的大规模高维时空数据, 具有长周期、大跨度等特点[1]; 对其运行模式的分析通常需要综合考量多类参数, 而对单个参数的分析结果往往不具有实际意义. 因此, 有必要从多参数分析角度对管道历史运行数据中蕴含的时空模式进行探索[2], 以有效地支持管道调度计划的合理安排及起泵方案的优化, 进而提高管道生产的经济效益[3]. 然而, 成品油管道数据分析领域中现有的支持多参数综合分析的方法以建模求解居多, 分析过程缺乏与领域知识紧密结合, 难以实现对长周期下历史多参数数据的综合时空模式进行交互探索[4]. 因此, 需要采用更灵活、有效的分析手段融合多参数并结合领域知识, 实现对多参数综合特征的时空模式提取与展示.

在时空模式交互提取方面, 通过设计复杂的时序视图与空间视图对时空数据进行可视化描述, 可以有效地挖掘大规模数据中存在的时空分布特征[5]; 但由于视图表述能力有限, 导致此类方法无法直观、高效地发现和提取时空模式. 通过特定的数据分析方法, 结合可视化视图也可以实现时空模式的提取. 金思辰等[6]通过异常识别等方法结合空间视图, 实现了对传染病数据的时空模式的交互式提取; 田东等[7]基于层次聚类等方法设计平行坐标等可视化视图, 探索了空气监测数据的时空特征及潜在规律. 但是, 由于分析过程与领域特定数据紧密关联, 导致此类时空模式提取方法的应用场景较为局限. 张量作为一种数学概念, 其本身多维、可扩展的特性可用于时空数据的组织, 对于时空数据建模的张量可以通过张量分解的方法提取时空模式. 罗月童等[8]基于张量分解的方法实现交通数据局部时空模式的探索分析. 但是, 普通的张量分解更多的是用于单参数时空模式的分析, 并不适用于成品油运行数据中多参数的应用场景.

为了解决上述问题及技术挑战, 本文结合可视化手段提出基于多参数融合的张量分解方法, 并设计了可视化系统 MPVis, 实现成品油管道运行多参数时空模式的有效提取及直观展示. 首先基于单参数的信息量及多参数间相关性, 结合用户领域知识进行参数分组融合; 然后将融合结果进行张量建模, 通过张量分解的方法提取其时空模式; 最后设计一组联动的多维模式视图和多参数对比视图, 支持用户交互式地探索多维时空模式, 并分析不同模式下多参数的具体变化趋势.

1 相关工作

1.1 成品油管道数据分析

随着智能管网的提出, 针对成品油管道数据的分析方法越来越多, 且大数据分析的手段已经在成品油管道数据分析领域受到关注, 但现有方法主要用于多参数建模求解方面. 针对管道调度计划的研究中, 张浩然等[9]通过对沿线站场进出站压力、流量、罐存体积等多类参数定义约束条件, 并建立混合整数非线性规划的数学模型求解, 实现了管道多批次油品输送的调度优化; 廖绮等[10]对进出站流量、进出站压力、密度及温度等测点数据融合建模, 并结合深度搜索算法求解, 实现调度计划的自动编制; Cafaro 等[11]提出一种基于离散时间模型的连续数学公式解决成品油管道调度, 提升了计算速度; Cafaro 等[12]提出成品油管道的数学规划模型, 通过对进出站流量、压力、密度等参数进行综合分析, 一次性确定最佳的输入和输出计划. 此外, 起泵方案的优化是成品油管道数据分析的另一个重点. 袁梦等[13]利用分支定界算法, 基于各站泵压力、进出站流量和油品密度等参数对管道沿线各站开泵的线性规划模型求解, 实现了成品油管道开泵方案统一优化模型. 虽然上述管道运行优化分析方法考虑了多参数的影响, 但使用的参数组合较为固定, 没有充分地利用历史数据中蕴含的时空规律实现更灵活的融合分析.

目前, 结合可视化技术的成品油管道数据分析方法在国内外还比较匮乏, 研究如何通过数据驱动的手段, 分析管道历史运行数据中多参数的时空模式, 并充分地利用可视化视图进行交互式探索分析, 具有重要的意义.

1.2 时空模式可视化探索

时空模式可视化探索的目的是发现和分析数

据在时空维度上的分布规律和变化趋势. 近年来，已有不少用于探索、提取及可视化数据中时空模式的方法. 贺一桐等[14]针对接警数据设计数据立方体，支持时空模式的交互式提取; Li 等[15]使用基于核密度估计的方法获取时空数据的分布规律，并实现 DDLVis 系统，用于交互式探索时空模式; Xiao 等[16]对工业级电力消费数据的时空相关性进行分析，设计开发了 EcoVis 系统，以支持用户交互式地分析不同行业用电模式的时空分布规律; Bai 等[17]面向轨迹数据提出矢量化的表示方法探索人类活动模式的时空规律，支持用户分析不同地区空间关系对人口移动模式的影响; Peña-Araya 等[18]通过设计柱状图、散点图叠加地图的可视化视图，描述时空相关性和时空模式; Wu 等[19]通过地图结合扇形图的方式综合可视化时空模式，配合直方图深入探索不同维度的时空模式的具体表现; Deng 等[20]通过可交互地图，结合折线图和日历图可视化空气质量数据中的异常时空特征: 上述时空模式提取及可视化的分析探索过程较为复杂，需要多次交互迭代提取时空模式，不能直观、高效地发现并总结数据中的时空特征，且大多是针对单一维度属性进行时空模式的探索分析，不具备多参数的分析能力. 通过张量组织数据并运用张量分解的方法，可以更加直接、有效地同时提取数据集在多个维度上的分布模式.

近年来，由于张量本身具备表述多维数据的能力，张量建模及张量分解的方法在时空数据分析中得到了广泛关注. 通常，张量分解包含 Tucker 分解和 CP(canonical polyadic)分解 2 种经典方法. Tucker 分解用于不同维度的相关性分析, Guo 等[21]利用张量对事件数据进行建模，并通过 Tucker 分解方法提取事件序列的时空相关性; Liu 等[22]利用基于 Tucker 的非负增广张量分解方法对交通数据进行分析，实现城市功能区划分的可视化探索. CP 分解[23]则主要用于获取不同维度下的独立特征模式，目前已用于分析各种时空数据，包括基于位置的社交网络数据[24]、自行车共享数据[25]和交通流量数据[26]. 在以上工作的基础上，Liu 等[27]进一步提出分段秩 1 张量分解的方法，用于大规模时空数据的渐进式子集划分及多维模式提取. 然而，上述工作利用张量分解方法提取时空模式时针对的是单个被分析参数构成的时空数据集，不能有效地支持成品油管道数据分析领域中面向多参数综合表征的时空模式提取.

本文将可视化设计与基于多参数融合的张量分解方法紧密结合，提出一种用于多参数时空模式交互提取与可视化探索的有效方法.

2　数据与任务分析

2.1　成品油管道数据

本文数据来源于某成品油管道历史运行数据，所用数据已去除涉密信息. 历史数据由集散控制系统提取自管道沿线各站场传感器，包括压力、流量、温度和密度等多种类别. 本文选用 45 d、8 个站场和 13 个参数共 1 300 万条数据进行实验，成品油管道典型监测参数描述如表 1 所示，由于成品油在管道中输运的特点，这些站场监测参数变化较为缓慢，时序敏感度较弱. 因此，对原始数据进一步取 20 min 为间隔然后求均值，最终得到 336 960 条数据.

表 1　成品油管道典型监测参数

字段名	意义	单位
STN_*_FI	进站流量	m^3/h
STN_*_PI1	进站压力 1	MPa
STN_*_PI2	进站压力 2	MPa
STN_*_DI	进站密度	kg/m^3
STN_*_TI	进站温度	℃
STN_*_FIC1	下载流量 1	t/h
STN_*_FIC2	下载流量 2	t/h
STN_*_FIQ1	累计下载流量 1	t
STN_*_FIQ2	累计下载流量 2	t
STN_*_LDI1	下载密度 1	kg/m^3
STN_*_LDI2	下载密度 2	kg/m^3
STN_*_PIB	减压阀前压力	MPa
STN_*_PIA	减压阀后压力	MPa

在进行成品油管道数据分析之前，需要对原始数据进行预处理. 原始数据包含日期、时间、站场、参数点位号和测量值，而不同参数之间的度量单位和取值范围可能存在差异. 为了确保数据的可比性和一致性，鉴于不同参数间度量单位不尽相同，须在统计分析前对所有控制参数进行 min-max 处理，数值统一归一化到[0,1].

2.2　任务分析

通过与成品油储运领域的专家进行反复交流和讨论，本文针对成品油管道运行中多参数数据时空模式，提出分析目标如下:

G1. 根据不同分析角度灵活划分参数组，融合多个参数特征.

G2. 提取和展示同组多参数综合特征的多维

时空模式.

G3. 支持分析综合特征的时空模式以及背后多参数的具体变化趋势.

基于上述分析目标，提炼出可视化及分析任务如下:

T1. 支持用户基于单参数信息熵和多参数相关性的度量，结合领域知识对监测参数进行不同角度的分组(G1). 由于使用单一角度的参数分析具有局限性，因此将具有相关性的多个参数组合成综合参数，从不同分析角度更全面地评价管道运行特征.

T2. 展示分组多参数融合过程，分析不同参数对融合结果的贡献程度(G1, G2). 融合后的综合参数蕴含的多维时空特征由多个原始参数共同作用，支持用户交互分析不同参数对综合参数特征的影响和贡献程度.

T3. 综合参数的多维时空模式提取与可视化表达(G2). 对综合参数进行张量建模和张量分解后，支持多参数综合特征在周期性、时段性和空间分布规律等多维时空模式上的有效表达和交互探索.

T4. 探索综合参数时空模式背后原始参数的具体变化趋势(G3). 综合参数的多维时空模式是对多个原始参数时空模式的融合，需要进一步将其与不同原始参数在时空维度上的具体变化趋势相关联，帮助用户发现更加详细的时空规律.

根据上述分析任务，数据处理和分析流程包括多参数评价、多角度分组融合、时空模式提取和多参数对比分析 4 部分，如图 1 所示.

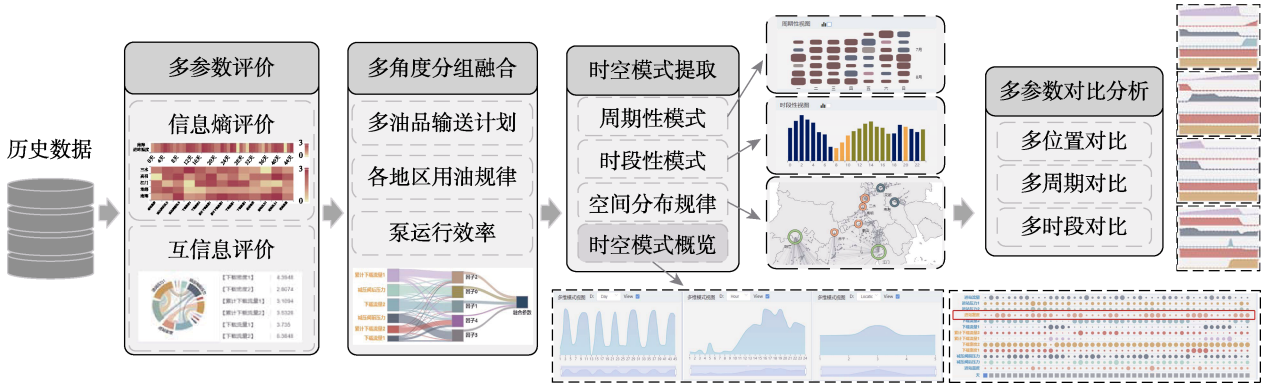


图 1 数据分析与处理流程

3 时空模式分析方法

3.1 多参数分组融合

本文采用信息论[28]中的信息熵和互信息这 2 种方法度量单参数信息量以及多参数间的相关性，以指导参数分组. 信息量可以帮助理解数据的不确定性和复杂度，提高数据分析的有效性；而互信息可以帮助发现数据之间的相关性和关联性，更好地理解数据之间的关系. 基于信息熵及互信息的方法对参数进行划分得到参数组，然后通过因子分析法对同组参数进行融合并计算参数贡献及权重，最终得到综合参数.

3.1.1 基于信息熵与互信息的参数分组

原始数据集包含 8 个站场、13 个监测参数的采样，不同站场、不同参数的数值变化及分布规律不同，表现出的数据复杂程度相差较大，通过计算参数信息熵，可以辅助用户更准确地评价原始数据集中不同参数的复杂程度. 对于参数的每个样本值，根据其在数据集中出现的频率计算其信息熵值，公式为

$$H(X)=E[I(X)]=E\left[\mathrm{lb}\frac{1}{P(x)}\right].$$

其中，$P(x)$ 表示不同参数值在数据集中出现的频率. 信息熵 $H(X)$ 的计算结果用于评价数据集中参数值的不确定性和复杂性，$H(X)$ 越大，表明该参数在样本集中数据取值越不稳定，数值差别较大，存在明显的动态变化特征；反之，说明该参数在样本集中越稳定，数值波动较小，动态特征不明显. 信息熵的评价方法可以帮助成品油储运领域研究人员从数值变化角度进一步识别与评价各个参数的分析价值.

互信息是一种评价参数间相关性的方法[29]，通过量化一个变量，并根据另一个变量数据获得的信息量，衡量 2 个随机变量之间的相互依赖关系；互信息越高，表明 2 个参数间的相关性越强，更适合将其划分到同一组中进行分析. 计算参数间的互信息时，首先根据每个参数的数值在数据集中的分布概率，基于式(1)计算其信息熵；然后

计算参数两两间的条件熵 $H(X|Y)$，即对于 2 个参数 X 和 Y，计算在给定 Y 的条件下 X 的条件熵，公式为

$$H(X|Y)=-\sum p(x,y)\mathrm{lb}\,p(x|y).$$

其中，$P(X,Y)$ 表示联合概率分布；$P(X|Y)$ 表示在已知 Y 的条件下，X 出现的概率，即条件概率分布．计算 2 个参数之间的互信息公式为

$$I(X:Y)=H(X)-H(X|Y).$$

互信息 $I(X:Y)$ 越大，表示 2 个参数间依赖性越强；$I(X:Y)=0$，表示 2 个参数相互独立．通过计算参数间的互信息评价参数相关性，相关性较强的参数可以划分为同组进行分析．

在实际分组过程中，首先通过信息熵判断参数的信息量，其值大时表示包含较多特征的参数；然后通过互信息判断参数间的相关性，将相关性较强的参数划分为同一参数组进行分析；在此基础上，结合领域知识和经验对参数组进行适当调整，达到更合理的参数分组．

3.1.2　基于因子分析法的参数融合

将原始数据集中的多参数划分为不同分析角度的多个参数组后，基于因子分析法[30]将同组多参数融合为一个表述多参数综合特征的综合参数．因子分析是一种对融合结果解释性较强的多参数融合方法．

首先基于多参数间的互信息构建参数相关性矩阵；然后根据 Kaiser 准则确定因子个数 N；再对特征向量进行计算，得出载荷矩阵和因子得分；最后将原始参数通过载荷矩阵和因子得分加权得到融合结果．其中，载荷矩阵表示不同参数对 N 个因子的贡献程度，若某参数对因子贡献总和最大，则对最终融合结果的影响也最大，可以将其作为综合参数的实际意义进行解释；因子得分反映了原始数据集在不同因子上的表现，得分较高的因子在综合参数中权重较大，此类因子中，贡献程度较大的参数可以作为综合参数的实际意义进行解释．基于因子分析法的多参数融合可以通过方差度量特征保留程度．

通过因子分析法对多参数的融合过程，可以识别不同参数对综合参数的影响程度，帮助用户分析并理解不同时空模式的具体意义．

3.2　时空模式提取

对原始多参数的分组融合可以得到描述综合参数的时空数据．对该时空数据集进行张量建模，并通过张量分解及聚类的方法提取其时空模式．综合参数张量建模方法流程如图 2 所示．

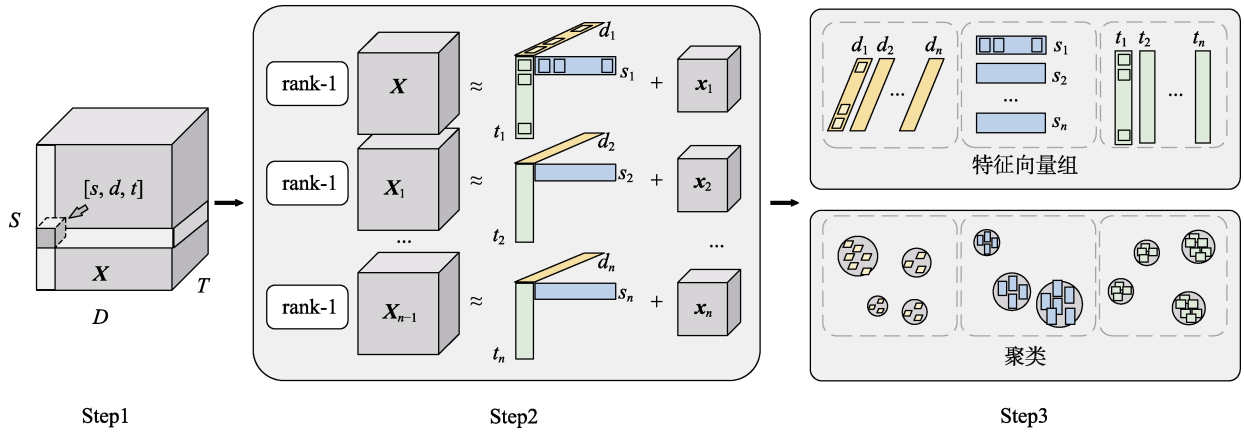


图 2　综合参数张量建模示意图

Step1．张量建模．将综合参数的时空数据建模为三阶张量 $S\times D\times T$．其中，S 表示多个站场，D 表示天数，T 表示时间段的个数；将每天再分成多个时间段，如按小时划分，得到的张量可表示为 $\boldsymbol{x}\in\mathbb{R}^{S\times D\times T}$，其中，$\boldsymbol{x}[s,d,t]$ 表示在站场 s 处第 d 天第 t 时间段综合参数的数值．

Step2．张量分解．将高阶张量分解为一组低维向量的积来近似拟合原始张量，分解后得到的低维向量中蕴含着原始张量在各维度上的变化趋势及分布情况．本文采用连续 rank-1 的 CP 分解方法，即每次分解时使用秩为 1 的张量来拟合原始张量；然后用原始张量减去拟合后的张量得到剩余张量，接着对剩余张量继续执行连续地秩为 1 的 CP 分解，直至达到设定的迭代次数或连续 2 次分解误差变化不大即可停止．其中，误差计算公式为

$$\begin{cases}\min\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|\\ \hat{\boldsymbol{x}}=\sum\boldsymbol{a}\odot\boldsymbol{b}\odot\boldsymbol{c}\end{cases}.$$

每次执行分解时，得到的一维向量组 (s,d,t) 都可以最大限度地保留原始张量的特征．本文将第 1 分解向量组

用来表示原始张量的多维时空模式概览，并对其进行可视化描述，从而得到综合参数在 S，D 和 T 维度上的变化趋势.

Step3. 模式提取. 上述步骤得到综合参数的多维时空模式不能很好地提取数据集中更丰富的时空特征. 通过连续秩 1 张量分解得到了多组低维向量，这些向量组都不同程度地保留了原始三阶张量的特征. 如图 3 所示，将每个张量中包含的 3 个一维向量按不同维度进行划分，得到描述不同维度特征的多维特征向量组；然后通过 *K*-means 聚类、层次聚类算法[31]和基于密度的 OPTICS 算法[32]多种聚类算法对不同维度的特征向量组进一步聚类，聚类结果可以展现出综合参数在不同维度上更丰富、更详细的时空模式.

根据不同的聚类结果，可以对不同时空模式下原始多参数的具体变化趋势进行进一步的对比分析，根据不同参数长周期下详细的变化趋势深入理解时空模式.

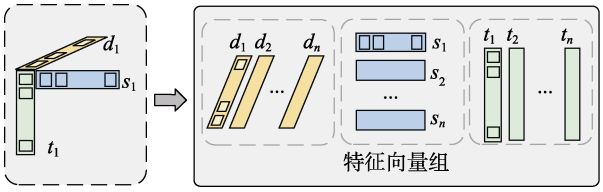


图 3 多维特征向量组

4 可视化设计

本文将数据分析方法与交互式可视化技术相结合，面向成品油管道运行多参数历史数据设计并实现了可视分析系统 MPVis，由 4 部分组成，其主界面如图 4 所示.

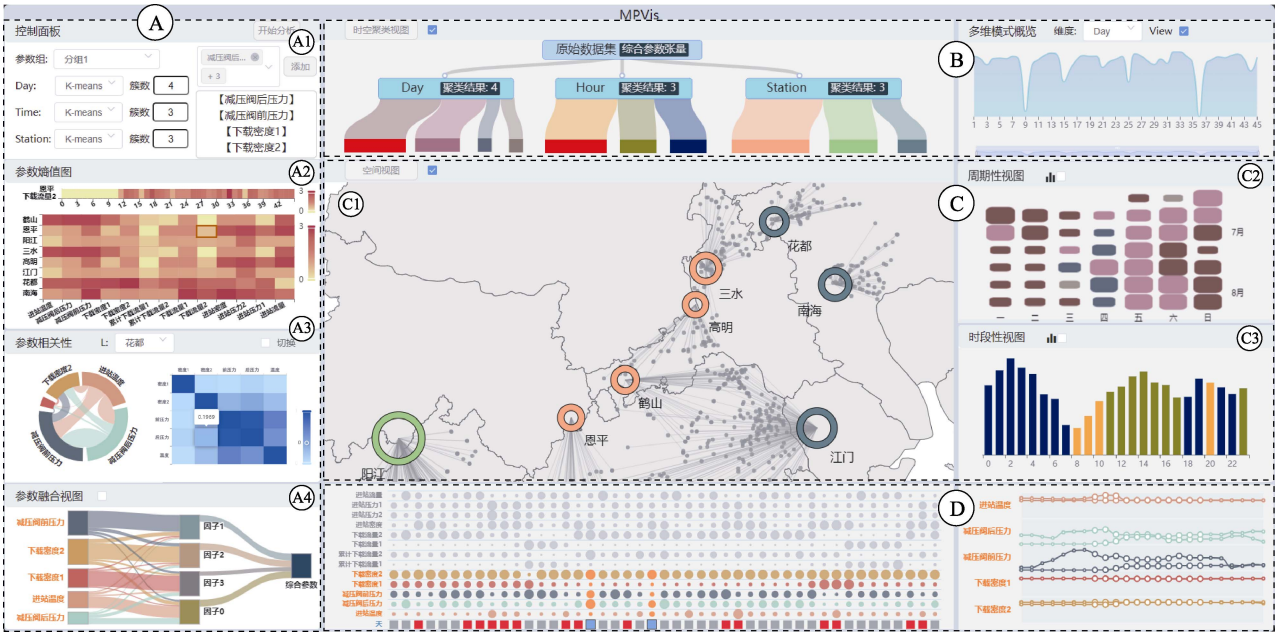


图 4 MPVis 系统主界面

4.1 参数分组融合视图

4.1.1 参数分组视图

根据第 3.1.1 节方法得到的信息熵和互信息，在图 4 的 A 区域 A2 和 A3 中展示每个参数的复杂程度及不同参数间的相关性(T1). A2 通过热力图来描述不同站场位置下原始参数信息熵的大小. 其中，每个矩形块表示不同站场、不同参数在 45d 内的信息熵均值. 此外，通过点击选择不同的块，将对应站场及参数的 45d 的详细信息熵展示到上面的条状热力图中. A3 左侧通过和弦图来描述原始参数间的相关性大小，不同参数使用不同的颜色进行标记，并将参数间互信息值的大小映射为弦的宽度.

参数相关性的定量描述如图 5 所示，其中，左侧的列表为每个参数与其他参数相关性的总和. 同时为了更准确地观察不同参数间互信息的具体数值，设计了右侧的矩阵热力图，用颜色深浅来映射不同参数间互信息的数值大小，当鼠标悬浮在某一矩形块上时就可以看到对 1 个、2 个参数间互信息的具体数值. 此外，在图 4 中 A3 标题右侧设计了用于切换视图的按钮.

4.1.2 参数融合视图

参数融合视图根据第 3.1.2 节的方法对同一参数组包含的多个参数进行融合，并对载荷矩阵及因子得分进行可视化表达(T2). 如图 4 中 A4 区域

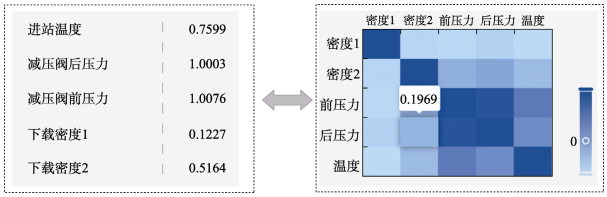


图 5　参数相关性视图

所示，通过桑基图展示不同参数对多个因子的贡献值，以及多个因子对最终综合参数的权重．参数的节点颜色与参数划分中的和弦图保持一致，桑基图中不同参数节点的高度表示参数对多个因子的贡献值总和，多个因子对融合结果的权重映射为因子与综合参数间连接线的宽度．通过原始参数对不同因子的贡献值以及因子在融合结果中的权重进行综合评价，可以判断不同原始参数对综合参数的影响程度．

4.2　时空模式分析视图

4.2.1　多维模式概览视图

基于第 3.2 节中连续秩 1 张量分解方法，得到的第 1 分解向量最大限度地保留了当前综合参数的时空特征，因此需要展示其在 *S*, *D* 和 *T* 多维属性上的变化趋势及分布情况(T3)．如图 4 中 B 区域右图所示，采用曲线图对不同维度的特征向量进行可视化，用户可以通过顶部选择框切换不同维度，观察当前综合参数在对应维度上的具体变化趋势及分布情况，为时空模式的探索分析提供概览．

4.2.2　聚类视图

为了进一步探索综合参数的多维时空模式，对第 3.2 节中不同维度模式提取得到的聚类结果通过桑基图结合树图来进行可视化(T3)．如图 4 中 B 区域左图所示，根节点为综合参数时空数据集构建的原始张量；第 1 层树图表示张量分解的不同维度；第 2 层通过桑基图表示不同维度下的聚类结果；其中，节点宽度表示此类别包含数据量的大小，如在 *D* 维度上，节点宽度表示此分类中包含天数的多少．聚类视图的节点颜色与第 4.2.3 节的时空模式视图对应，且支持用户交互选择不同类别，选择结果关联到空间视图及多参数对比视图中进行高亮显示．

4.2.3　时空模式视图

在多维模式概览及聚类视图的基础上，通过周期性视图、时段性视图和空间视图对不同维度聚类结果进行详细可视化(T3)，如图 4 中 C 区域所示．

(1) 周期性视图．对 *D* 维度的特征向量组聚类结果描述的特征表现在周期性模式上．在图 4 中 C2 区域的日历图中，不同块的大小映射为综合参数的特征数值，块的颜色映射为不同类别，通过日历图可以直观地表达出综合参数在不同周上的周期性模式．如图 6 所示，日历图还可以切换为柱状图，以清晰地对比不同 *d* 的数值差别大小，进而发现数值在 *D* 维度的周期性波动．柱状图的颜色映射为聚类结果的不同类别，与聚类视图中 *D* 维度的聚类颜色保持一致．

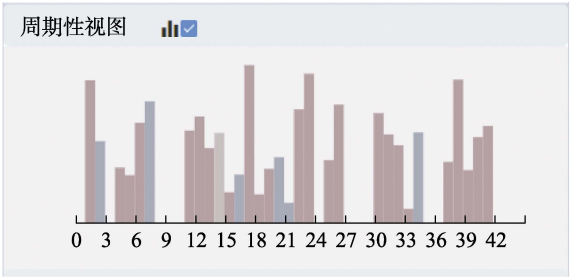


图 6　周期性视图

(2) 时段性视图．对 *T* 维度特征向量组聚类结果的特征表现在时段性模式上．柱状图可以有效地展示多个类别数据之间的对比，因此采用柱状图可视化 *T* 维度的详细聚类结果．如图 4 中 C3 区域所示，柱状图中，颜色对应不同类别并与聚类视图中时间段维度聚类颜色保持一致，高度对应综合参数在不同时段的数值大小，使用户更容易发现综合参数在 *T* 维度不同时段的差异．

(3) 空间视图．对 *S* 维度特征向量组的聚类结果是将站场分为不同的类．空间位置最直观的可视化方式是使用地图，但需要借助其他图元来表示聚类结果及数值大小．如图 4 中 C1 区域所示，本文通过地图可视化各个站场在空间位置上的分布，并通过站场中心位置处气泡的大小映射其特征值，颜色用来映射不同类并与聚类视图中 *S* 维度聚类颜色保持一致．通过图元大小差异及颜色区分，可以直观地识别综合参数在空间维度上的分布模式．

4.3　多参数对比视图

为了进一步挖掘并分析综合参数时空模式背后的原因，本文对综合参数对应的原始多参数组中的数据进行关联可视化(T4)，实现对数据的具体变化趋势的分析．如图 4 中 D 区域所示，左图为多参数不同日期中均值的对比视图，纵轴对应多个参数，当前参数组包含的参数高亮显示，横轴对应不同的“天”，每个参数行内点的大小映射为原始数据一天内的均值；右图为参数一天内的变化趋势，通过折线图来描述，当聚类视图选中某类时，该类

别包含时段对应的点高亮显示；点击左图底部可以选择多个日期进行观察对比，选择的“天”对应的矩形块变为蓝色，对应列的散点变为橘黄色，同时右图更新为对应多个日期的参数变化趋势.

为了实现原始多参数的多维可视化，本文设计了聚类视图、空间视图与多参数对比视图之间的联动功能. 在图 4 中 B 区域的聚类视图中，用户可以点击 D 和 T 维度的节点，指定对应的不同日期和时段. 在图 4 中 C1 区域的空间视图中，用户可以选择不同站场，多参数对比视图随即更新；图 4 中 D 区域的左图底部选择框对应的不同日期会变为红色，右图对应不同时段的点会高亮显示. MPVis 系统采用多视图联动交互设计展示挖掘出的时空模式，可以同时支持对不同维度模式的对比以及多参数变化趋势的对比分析；此外，结合数据分组过程可以支持用户更好地理解多维时空模式，降低时空模式探索过程的复杂性.

5 案例分析及专家评估

通常，成品油管道由多个输油站场组成，根据不同地区用油需求将成品油从炼油厂输送至沿途各地区，完成用油供给. 成品油管道数据分析中主要关注对输油规律的挖掘，以支持生产及调度计划的制定. 本文根据第 2.1 节描述的某成品油运输管道历史运行数据，从管线输运模式和局部地区用油模式 2 个具有实际意义的角度进行案例分析；邀请 2 位该领域的专家深度参与系统试用和评估，其中，1 位是油气储运调度优化研究人员，另外 1 位是管道公司现场运行人员.

5.1 管线输运模式分析

管线输运模式分析指对长距离管道输送过程中监测参数的模式特征进行分析，通过对成品油管线输运模式的分析，可以支持不同油品输运周期的安排，优化多油品混合输运方案，为优化成品油输送计划提供有力的依据和方向.

图 7 所示为对实际输运情况的时空模式进行了探索和分析的案例. 该案例结合现场运行人员的经验知识，将下载密度、累计下载流量和进站密度 3 类参数划分为同一参数组，综合评价实际输运情况的时空模式. 实验结果表明，实际输送情况在 D 维度上呈现明显的周期性规律，且周期长度大致为一周，而在空间维度上存在特殊站点，表明此站点输运模式不同于其他站点.

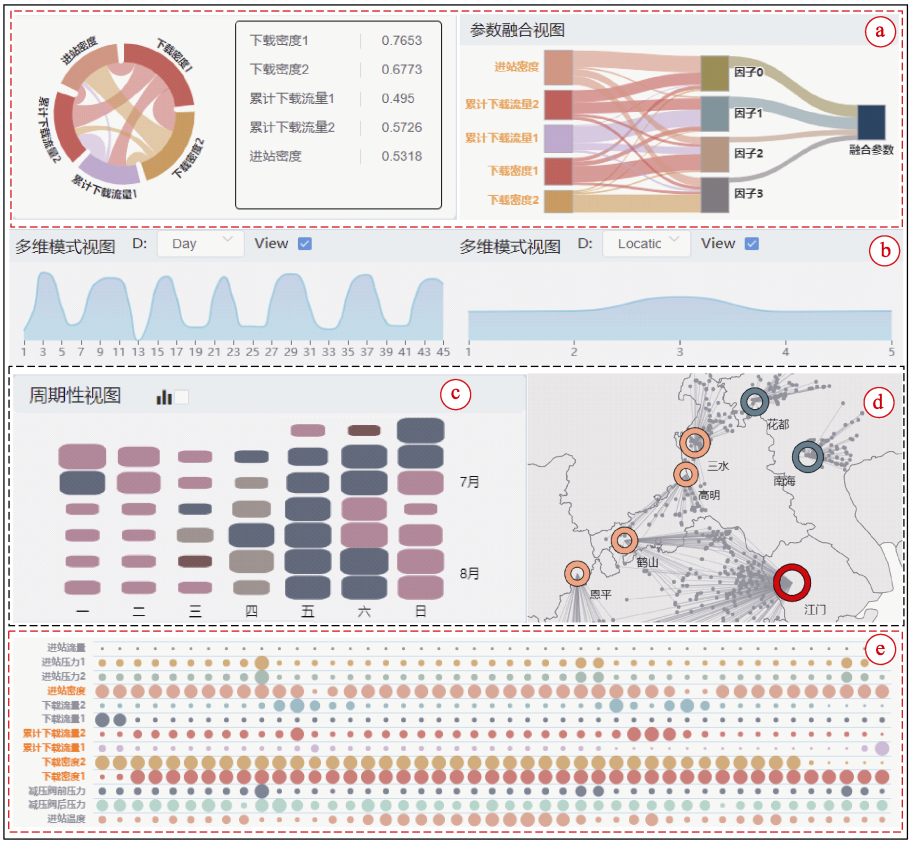


图 7 管线输运模式分析

从图 7 可以看出：

(1) 区域 c 中，在 D 维度上将 45 d 聚成 3 类，通过参数融合视图发现，因子 1 对综合参数贡献较大，而进站密度参数在因子 1 中构成占比最大，因此进站密度对最终综合参数的影响较大，而进站密度表示不同油品种类，所以该成品油管道实际输送中包含 3 种不同的油品，且油品密度相差较大. 由该图得出，综合参数存在周期性规律，在输油过程中通常会对 3 种不同密度油品进行固定顺序输送，输送模式呈现周期性规律.

(2) 区域 c 的周期性视图中，通常在每周的前半周综合参数值较小，后半周则较大，具有以周为间隔的周期性模式. 区域 a 的参数融合视图中，因为密度参数对综合参数贡献较大，所以用油品密度解释综合参数，而高密度油品通常是#0 柴油，低密度油品通常是#92 汽油和#95 汽油. 领域专家指出，该管线主要为珠三角地区供油，由于周中工作日出行较多，城市地区对#92 汽油和#95 汽油需求增大，导致管线前半周输送低密度油品较多；因此管道的输运模式存在以周为间隔的周期性模式.

(3) 区域 d 的空间视图中，灰色小点为该站场附近的加油站，在空间维度模式上表现特殊的站点(江门)附近加油站相对较多，因此，管线的输运模式在空间维度受不同站场供给加油站多少的影响.

本案例中，领域专家通过 MPVis 系统提取的时空模式，可以进一步得出输油过程在不同站场及不同时间周期下的一些普遍规律及特征，帮助调度优化研究人员更多地了解管道在长周期下的输油规律，获得更多的输运模式特征，支持整体输油过程的安排和优化.

5.2　局部地区用油模式分析

局部地区用油模式分析对成品油输送尤为重要，对管线沿途地区用油情况进行分析可以发现用油规律，揭示不同地区的用油特征及潜在的用油原因，从而更精准地制定油品下载计划，为优化储油方案提供参考依据.

图 8 所示为领域专家基于参数相关性，将进站压力、下载流量、累计下载流量和下载密度 4 类参数划分为同一参数组，评价地区用油情况的案例. 时空模式提取结果在空间维度上聚成 3 类，其中一类只包含 1 个站场(阳江)，如图 8 区域 b 所示，表明此站场附近地区用油模式较为特殊，有必要对其进一步分析.

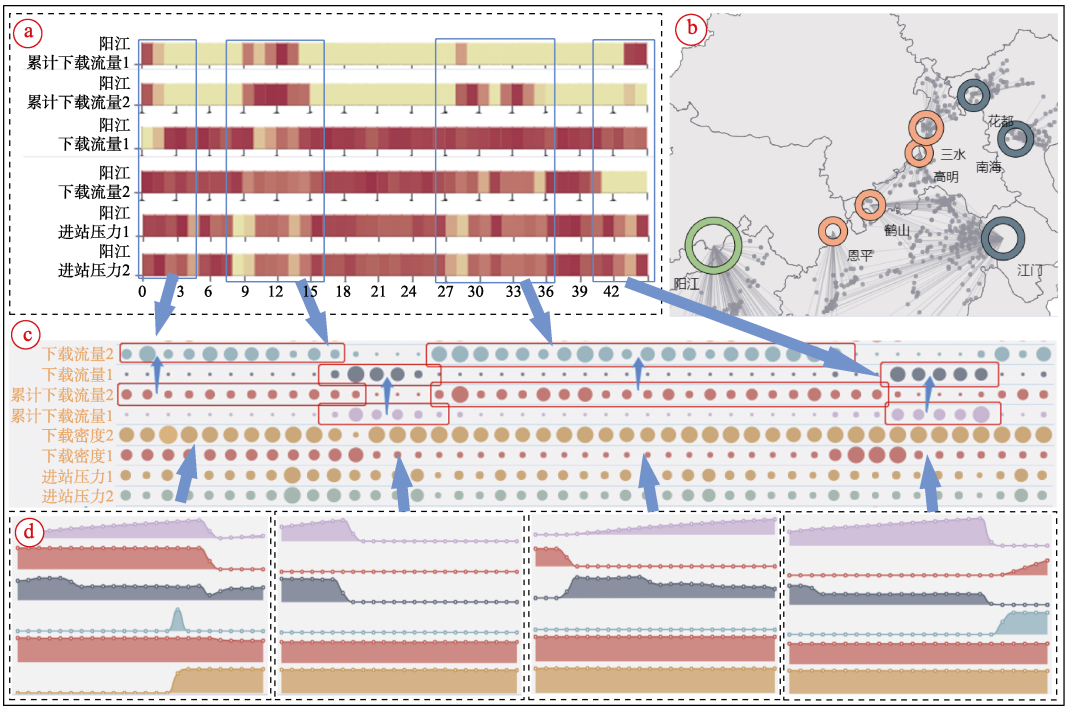


图 8　局部地区用油模式分析

通过参数熵值视图和多参数对比视图，对此站场多参数 45 d 周期下的变化趋势进行对比分析，发现该地区用油模式如下：

(1) 如图 8 区域 a 所示，在评价局部用油模式的多参数中，进站压力、下载流量和累积下载流量信息熵的变化趋势在 45 d 内呈现一致的对应关系，

表明此 3 类参数具有较强相关性，且稳定性特征表现一致，表征了一种较为明显的模式. 现场运行人员指出，在 45 d 周期的不同时间段内，下载流量和累计下载流量表征该站场对不同种类油品的输运情况，调度优化研究人员进一步基于工艺流程判断在 45 d 周期内存在 2 类参数对应关系，分别为汽油和柴油.

(2) 通过选择参数突变的日期，可以在多周期、多时段对比视图中具体查看各参数的数值表现，如图 8 区域 c 和区域 d 所示，可以看出，多参数的变化趋势与参数熵值图呈明显对应关系，模式表征一致. 进一步，现场运行人员根据图 8 中 c 区域的下载密度分布得出，此站场在 45 d 中，柴油下载量较大的天数有 35 d，汽油需求较大的天数有 10 d. 专家推测，因为该站场所在地区重工业较多，工业用油较多，所以用油模式与其他地区有明显区别.

(3) 通过对空间维度聚类结果中 3 类站场时空模式对比分析得出，图 8 区域 b 中橘色类别的 4 个站场为一类，其主要用油模式为城市用油；墨绿色类别的 3 个站场附近有少量工业厂区，但其用油模式仍以城市用油为主；浅绿色类则为工业用油模式. 不同地区周边环境不同，导致用油模式区别较大.

在此案例中，领域专家通过 MPVis 系统提取的时空模式挖掘出更多局部地区用油量及用油种类的规律，了解了不同地区的储油能力及用油方案，进一步帮助领域专家制定下载计划及储油安排.

5.3 算法定量评估

为了评估本文方法在张量分解过程中对多参数特征的保留程度，参考文献[27]构造了多参数时空模式提取的基线方法，并利用上述 2 个案例数据，从不同时空维度分别对 2 种方法的多参数特征损失误差进行计算，对比分析本文方法的有效性，其中，多参数特征损失误差越小越好.

为了提取多参数的综合特征，在构造的基线方法中，首先将前文分析的多参数分别进行张量建模并执行连续秩 1 分解，得到每个参数对应的多维特征向量组；然后针对不同时空维度将多个参数的特征向量组进行整合，得到 L 个大小为 $\boldsymbol{N}\times\boldsymbol{P}$ 的向量. 其中，L 表示某一时空维度的长度(如天数)，N 表示连续秩 1 分解的迭代次数，P 表示参数数量. 在对此向量组进行聚类的基础上，根据不同类别将每个参数对应的张量划分为多个子张量；最后对子张量执行秩 1 分解，得到反映综合特征的第 1 分解向量组，计算不同参数的特征损失误差并求平均值，得到提取多参数综合特征的整体误差水平. 而在本文方法中，首先根据综合参数连续秩 1 分解后构建的特征向量组进行聚类；然后根据聚类结果将综合参数对应的张量划分为子张量；再对子张量执行秩 1 分解；最后相对于原始参数，分别计算综合参数的损失误差并求均值，得到相应的整体误差水平.

本文方法和基线方法的对比结果如表 2 所示. 可以看出，本文方法在各个维度上都有更好的表现；基线方法在聚类前对特征向量组的直接整合使得描述某一时空维度的特征过多造成聚类效果不佳，导致在后续的张量分解中特征损失误差较大. 实验结果表明，本文方法可以更好地保留多参数的综合时空特征，为分析人员提取到更多的时空模式，同时提取的结果更可信.

表 2 2 种方法特征损失对比

方法	维度					
	站场		日期		时段	
	案例 1	案例 2	案例 1	案例 2	案例 1	案例 2
基线	11 680	13 550	6 818	7 396	9 750	14 362
本文	3 835	4 864	676	773	4 689	3 694

5.4 专家反馈

MPVis 系统设计与开发过程中始终与领域专家保持紧密合作，经过反复修改，历时一年多时间最终形成目前的原型系统. 为了更全面地了解领域专家对该系统的意见和建议，进一步邀请上述调度优化研究人员与现场运行人员对系统的有效性和可用性进行评估和反馈.

调度优化研究人员认为，MPVis 系统直观交互流畅，可以快速地代入领域经验深入分析多个站场历史数据中蕴含的输运规律，有效地支持了生产任务制定过程的优化，具有较强的实用价值. 现场运行人员指出，本文方法及可视化手段能够在很大程度上降低工作复杂度，对现阶段成品油管道数据分析领域基于数据驱动的决策支持具有较大的指导意义.

上述领域专家对 MPVis 系统呈现的优势和进一步改进方面提出以下反馈:

(1) 交互分析方面. 专家们认可现有交互分析流程，认为分析逻辑严密，能够支持每一步的推理论证. 在实际操作过程中，他们也能灵活地从各个视图中找到想要的信息，同时能将自己的专业知识与交互过程紧密结合. 不足之处在于，当前的时

段多参数对比视图可以支持不同日期之间的比较，但选择的天数较多时会导致视图比较杂乱，建议对此部分功能继续优化.

(2) 可视化设计方面. 领域专家对系统视图整体设计非常满意，与现在油气领域数据展示方式相比有了很大改进. 现场运行人员希望在空间视图中可以包含更多的信息，如经济、交通等数据，更加便捷地获取管线所在地区的更多参考信息.

(3) 扩展性方面. 该系统可以从不同分析视角和不同时空范围对成品油管道运行参数进行灵活组合，支持提升管道生产经济性和安全性为目标的多种综合时空模式探索. 为了适应分析更大规模和更高维度的管道时空数据，MPVis 系统在数据处理能力和系统性能方面还需进一步优化.

6 结　语

面向成品油管道历史运行多参数数据，本文提出一种基于多参数融合的张量分解方法，用于提取数据中蕴含的时空模式. 从多参数分组融合、时空模式提取及可视化、原始多参数对比分析等角度设计了可视化系统 MPVis. 首先通过分析单参数信息量及多参数相关性，支持用户结合领域知识对多参数进行分组并选定参数组进行分析；然后对选定参数组进行参数融合，实现多维时空模式的提取及可视化；最后对时空模式进行交互式探索，以及对不同模式下多参数的原始数据变化趋势进行对比分析，支持用户深入理解多参数综合表征的时空特征及分布规律. 通过真实案例数据进行分析，验证了 MPVis 系统的可用性和有效性. 虽然本文工作针对的是成品油管道运行数据，但从分析方法和可视化设计上看，MPVis 系统可推广到其他更广泛的工业领域，为多参数综合表征的时空模式提取与分析提供有力支持.

参考文献(References):

[1] Zhou Xingyuan, Liang Yongtu, Zhang Xin, *et al*. Research progress of reliability analysis of product oil pipeline network[J]. Oil & Gas Storage and Transportation, 2021, 40(6): 629-636+691(in Chinese)
(周星远, 梁永图, 张昕, 等. 成品油管网可靠性分析的研究进展[J]. 油气储运, 2021, 40(6): 629-636+691)

[2] Khlebnikova E, Sundar K, Zlotnik A, *et al*. Optimal economic operation of liquid petroleum products pipeline systems[J]. AIChE Journal, 2021, 67(4): Article No.e17124

[3] Shen Yun. Research progress on the operation optimization of multiproduct pipelines[J] Chemical Enterprise Management 2021(7): 152-154+181(in Chinese)
(沈允. 成品油管道运行优化的研究进展[J]. 化工管理, 2021(7): 152-154+181)

[4] Wang Z F, Li S Z. Data-driven risk assessment on urban pipeline network based on a cluster model[J]. Reliability Engineering & System Safety, 2020, 196: Article No.106781

[5] Sun D, Huang R F, Chen Y Z, *et al*. PlanningVis: a visual analytics approach to production planning in smart factories[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(1): 579-589

[6] Jin Sichen, Tao Yubo, Yan Yuyu, *et al*. Infectious disease patterns analysis based on visualization of multidimensional space-time data[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(2): 241-255(in Chinese)
(金思辰, 陶煜波, 严宇宇, 等. 基于多维时空数据可视化的传染病模式分析[J]. 计算机辅助设计与图形学学报, 2019, 31(2): 241-255)

[7] Tian Dong, Li Guan, Cheng Shiyu, *et al*. Visual analysis system for fine-grained inline relationship of air quality data[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(9): 1326-1336(in Chinese)
(田东, 李观, 程世宇, 等. 空气质量细粒度数据内联关系可视分析系统[J]. 计算机辅助设计与图形学学报, 2021, 33(9): 1326-1336)

[8] Luo Yuetong, Liu Xinyue, Guo Zhengyue, *et al*. Tensor decomposition based visual analytics method of traffic accident local spatio-temporal patterns[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(8): 1222-1233(in Chinese)
(罗月童, 刘新月, 郭正跃, 等. 交通事故局部时空模式的张量分解可视分析方法[J]. 计算机辅助设计与图形学学报, 2021, 33(8): 1222-1233)

[9] Zhang Haoran, Liang Yongtu, Wang Ning, *et al*. Optimal scheduling of multi-source single-distribution pipeline with multi-batch sequential transportation[J]. Acta Petrolei Sinica, 2015, 36(9): 1148-1155(in Chinese)
(张浩然, 梁永图, 王宁, 等. 多源单汇多批次顺序输送管道调度优化[J]. 石油学报, 2015, 36(9): 1148-1155)

[10] Liao Qi, Liang Yongtu, Zhang Haoran, *et al*. An automatic scheduling method for product pipelines based on depth first search method[J]. Oil & Gas Storage and Transportation, 2017, 36: 1391-1400(in Chinese)
(廖绮, 梁永图, 张浩然, 等. 基于深度搜索法的成品油管道调度计划自动编制[J]. 油气储运, 2017, 36(12): 1391-1400)

[11] Cafaro D C, Cerdá J. Rigorous scheduling of mesh-structure refined petroleum pipeline networks[J]. Computers & Chemical Engineering, 2012, 38: 185-203

[12] Cafaro V G, Cafaro D C, Méndez C A, *et al*. MINLP model for the detailed scheduling of refined products pipelines with flow rate dependent pumping costs[J]. Computers & Chemical Engineering, 2015, 72: 210-221

[13] Yuan Meng, Liang Yongtu, Zhou Xingyuan, *et al*. A unified model for optimizing the pump start-up scheme of products pipeline[J]. Oil & Gas Storage and Transportation, 2019, 38(9): 996-1002(in Chinese)
(袁梦, 梁永图, 周星远, 等. 成品油管道开泵方案统一优化

模型[J]. 油气储运, 2019, 38(9): 996-1002)

[14] He Yitong, Li Jie, Tao Wenyuan, *et al*. Interactive exploration of semantic-spatial-temporal patterns of large-scale alarm police logs[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(4): 518-531(in Chinese)
(贺一桐, 李杰, 陶文源, 等. 大规模接警日志语义时空模式交互探索[J]. 计算机辅助设计与图形学学报, 2021, 33(4): 518-531)

[15] Li C H, Baciu G, Wang Y Z, *et al*. Ddlvis: real-time visual query of spatiotemporal data distribution via density dictionary learning[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 28(1): 1062-1072

[16] Xiao Y, Zheng K H, Lonapalawong S, *et al*. EcoVis: visual analysis of industrial-level spatio-temporal correlations in electricity consumption[J]. Frontiers of Computer Science, 2022, 16(2): Article No.162604

[17] Bai J H, Zhang H J, Qu D Z, *et al*. FGVis: visual analytics of human mobility patterns and urban areas based on F-GloVe[J]. Journal of Visualization, 2021, 24(6): 1319-1335

[18] Peña-Araya V, Pietriga E, Bezerianos A. A comparison of visualizations for identifying correlation over space and time[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 26(1): 375-385

[19] Wu Y C, Weng D, Deng Z K, *et al*. Towards better detection and analysis of massive spatiotemporal co-occurrence patterns[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3387-3402

[20] Deng Z K, Weng D, Chen J H, *et al*. AirVis: visual analytics of air pollution propagation[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(1): 800-810

[21] Guo S N, Xu K, Zhao R W, *et al*. Eventthread: visual summarization and stage analysis of event sequence data[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(1): 56-65

[22] Liu L Y, Zhang H X, Liu J X, *et al*. Visual exploration of urban functional zones based on augmented nonnegative tensor factorization[J]. Journal of Visualization, 2021, 24(2): 331-347

[23] Carroll J D, Chang J J. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition[J]. Psychometrika, 1970, 35(3): 283-319

[24] Han Y F, Moutarde F. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization[J]. International Journal of Intelligent Transportation Systems Research, 2016, 14(1): 36-49

[25] Yan Y Y, Tao Y B, Xu J, *et al*. Visual analytics of bike-sharing data based on tensor factorization[J]. Journal of Visualization, 2018, 21(3): 495-509

[26] Chen X Y, He Z C, Wang J W. Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition[J]. Transportation Research Part C: Emerging Technologies, 2018, 86: 59-77

[27] Liu D Y, Xu P P, Ren L. TPFlow: progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(1): 1-11

[28] Shannon C E. A mathematical theory of communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2001, 5(1): 3-55

[29] Liu X T, Shen H W. Association analysis for visual exploration of multivariate scientific data sets[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 955-964

[30] Yong A G, Pearce S. A beginner’s guide to factor analysis: focusing on exploratory factor analysis[J]. Tutorials in Quantitative Methods for Psychology, 2013, 9(2): 79-94

[31] Jain A K, Murty M N, Flynn P J. Data clustering: a review[J]. ACM Computing Surveys, 1999, 31(3): 264-323

[32] Ankerst M, Breunig M M, Kriegel H P, *et al*. OPTICS: ordering points to identify the clustering structure[J]. ACM SIGMOD Record, 1999, 28(2): 49-60