

智能家居产品链服务平台的数据中台架构研究

钱苏杭, 顾 瑞, 付翔宇, 肖 健, 胡 静, 李 行

天津大学智能与计算学部, 天津

收稿日期: 2024年12月16日; 录用日期: 2025年1月8日; 发布日期: 2025年1月20日

摘 要

随着智能家居的广泛应用, 为应对智能家居系统在数据管理上的挑战, 文章提出并设计实现了一站式数据资源管理服务——智能家居产品链服务平台数据中台。该数据中台涵盖数据架构、数据标准、数据质量及数据生命周期管理等关键功能, 确保业务应用能够获取全量、标准化、干净和智能的数据。本平台主要由接口服务层、数据与算法服务层、大数据能力层以及云计算能力层四大部分组成。平台提出并设计实现了关键功能, 包括AI算法服务、智能画像分析和全景监控等。并且通过实现标准数据模型管理、应用数据模型管理、数据质量管理、数据链路分析以及智能标签体系建设等, 提供了高标准、易用的一站式大数据管理解决方案, 能够增强企业数字化管理能力, 为企业提供高质量的标准数据模型, 显著减少重复开发的工作量。同时, 能够全景、全链路地掌握数据质量、使用情况和系统运行状态, 从业务视角更直观地使用和探索数据, 从而更高效地挖掘和利用数据中的业务价值。

关键词

智能家居, 服务平台, 数据中台, 数据管理

Research on the Data Middle Platform Architecture of Smart Home Product Chain Service Platform

Suhang Qian, Rui Gu, Xiangyu Fu, Jian Xiao, Jing Hu, Hang Li

College of Intelligence and Computing, Tianjin University, Tianjin

Received: Dec. 16th, 2024; accepted: Jan. 8th, 2025; published: Jan. 20th, 2025

Abstract

With the wide application of smart homes, in order to cope with the challenge of data management in smart home systems, this paper proposes and designs a one-stop data resource management service—smart home product chain service platform data center. The data center covers key functions

文章引用: 钱苏杭, 顾瑞, 付翔宇, 肖健, 胡静, 李行. 智能家居产品链服务平台的数据中台架构研究[J]. 图像与信号处理, 2025, 14(1): 85-91. DOI: 10.12677/jisp.2025.141008

such as data architecture, data standards, data quality, and data lifecycle management to ensure that business applications can obtain full, standardized, clean, and intelligent data. This platform is mainly composed of the interface service layer, data and algorithm service layer, big data capability layer, and cloud computing capability layer. The key functions of the platform are proposed and designed, including AI algorithm service, intelligent portrait analysis, and panoramic monitoring. By implementing standard data model management, application data model management, data quality management, data link analysis, and intelligent label system construction, the platform provides a high-standard and easy-to-use one-stop big data management solution, which can enhance the digital management capability of enterprises, provide enterprises with high-quality standardized data models, and significantly reduce the workload of repeated development. At the same time, it can grasp the data quality, usage, and system running status in a panoramic and full-link way and use and explore data more intuitively from the business perspective so as to mine and utilize the business value in the data more efficiently.

Keywords

Smart Home, Service Platform, Data Middle Platform, Data Management

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景介绍与相关工作

随着智能家居技术的迅猛发展,智能家居在提升用户生活便利性和舒适性的同时,也带来了前所未有的数据管理挑战。智能家居生态系统开放性不足,抗风险能力差,生成的大量数据不仅种类繁多,而且具有高频率和高复杂性;并且智能家居企业水平参差不齐,协同效率低下,如何高效地管理和利用这些数据成为行业发展的关键。传统的智能家居数据管理系统往往面临数据管理上的诸多难题,包括数据整合不畅、数据质量难以保障、数据标准不统一等问题,这些问题严重影响了业务应用的效果和数据价值的发挥[1]。

在传统产业的数字化转型进程中,国内外已有相关的用于辅助产业链的数字化改革的互联网服务平台的研究,王昀[2]等人提出了数字政府的时空数据中台设计,并提供数据存储、数据计算和数据共享等功能,将数据接入城市大脑场景中,统筹管理各类时空上的数据资源,解决了数据分散的问题,并提高了数据利用率;Liu [3]等人提出了 DIGICOR 数字平台的架构设计,在工业 4.0 的环境下为中小企业提供开放性、模块化的治理机制和策略,解决了中小企业在数字化转型过程中平台支持上的缺失问题,提升了中小企业的协作能力和产业链管理能力;Christou [4]等人提出了应用于生产车间数据管理的 INTRA-QM 数字平台架构,集成了数据互操作、数据分析和数据源整合等功能,为产业的数字化管理提供可配置、可编辑和可互操作的解决方案,并在多个工业环境中进行了部署,证明了其有效性[5]。

为应对挑战,本文提出了智能家居产品链服务数据中台。本平台的核心目标是提供一站式的数据资源管理服务,涵盖数据架构、数据标准、数据质量控制以及数据生命周期管理等关键领域。主要通过数据处理全场景监控与算法控制等技术,通过创新性地引入人工智能算法与数据血缘等技术,设计并实现一个标准化程度高、易用性强的一站式大数据管理平台。从而帮助企业设计高质量的标准化数据模型,减少重复开发工作,还使企业能够全面掌握数据质量、使用情况和系统运行状态,从而更高效地挖掘和利用数据中的业务价值。

2. 平台设计与实现

智能家居产品链服务平台的数据中台需涵盖多个关键技术与服务领域，以确保全面、高效的数据管理。本章将介绍三项核心技术的设计与实现：基于语义与机器学习的资源发布式索引构建技术，通过深度学习构建词向量库，实现高效资源查询；业务驱动的资源要素资源服务化匹配技术，结合功能与性能匹配框架，提高资源匹配的效率和精确度；数据驱动的多维度用户画像技术，综合运用用户交互数据，构建个性化智能家居服务平台，如图 1 所示。

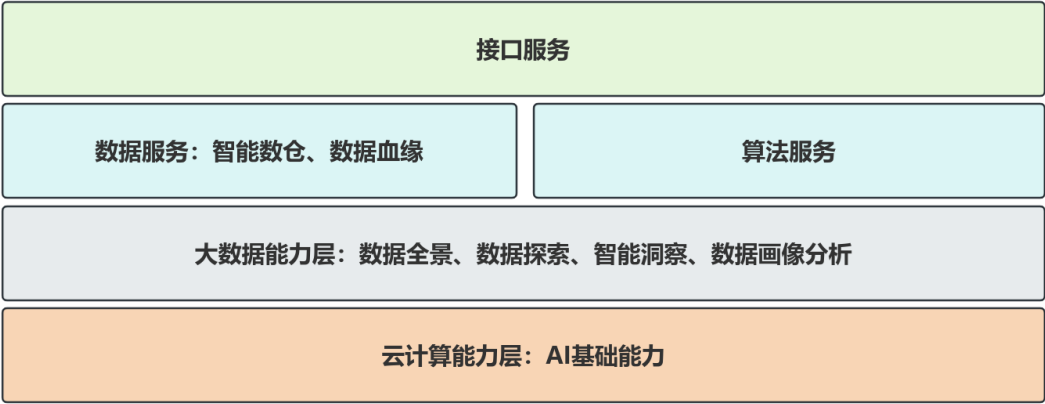


Figure 1. Data center overview map
图 1. 数据中台概览图

2.1. 基于语义与机器学习的资源发布式索引构建技术

本文面向供应链协同的主数据，利用深度学习技术构建专业的词向量库，并基于语义和词向量建立数据本体模型。同时，通过分类和聚类工具构建高可用、可扩展的数据索引存储，以提升资源查询的效率。

在处理产业链协同体系中的海量资源时，我们在资源索引构建中引入了语义分类方法，使得传统的 IR 树具备语义特征，从而支持基于语义的资源查询[6]。同时，提出了一种 Top-k 语义查询优化算法，并通过实验分析评估了该算法在实际数据中的有效性，结果表明该算法在索引构建时间、查询时间和查询准确率等方面表现良好。

对于一个特定的资源库 D ，根据资源分类体系，将资源要素编码并分为 7 类，参照经典的索引做法，对于第 i 类资源的资源个体，用一个二元组将其表示为 $p: f$ ，其中， p 表示资源个体的位置信息， f 表示资源个体的特征信息。对于一个典型的业务查询 q ， $p: f: r$ 分别表示查询的资源个体的位置信息、特征描述，以及适当空间范围查询半径。

本文提出一个基于语义分类的资源要素索引(SIR-Tree)构建方法，算法的伪代码如表 1 所示，算法的输入为待插入的对象集合 O ，以及待插入树的根节点(如果是第一个插入的对象，将该节点赋值为根节点)。算法的输出为插入了新对象的语义 IR 树的根。

算法主要流程如下：首先，寻找离待插入点最近的节点，如果该节点不是叶节点，则递归调用插入函数，反之，则对该节点进行插入操作，见算法的 1~5 行，这些操作并没有涉及到语义关联的内容，符合分离目标 1，因为资源的地理位置仍然是考虑的重点，如果两个地方相隔太远，那么无论有多高的相似度，资源都不能挂在同一父叶节点下。在算法的 6~9 行，如果节点下面的子节点数量大于预定的 M 值，则调用分离函数，如果没有，则将对象挂在叶节点上，同时更新叶节点的空间范围。如算法的 10~13 行。

Table 1. Semantic-based resource element index construction method
表 1. 基于语义的资源要素索引构建方法

算法 1. 基于语义的资源要素索引构建方法
输入：待插入的对象集合 O ，待插入树的根节点 $root$
输出：插入了新对象的语义 IR 树的根 $root$

1. 寻找初始待插入节点 $O_i \in O$
2. Node Insert (node, O_i)//递归调用插入函数完成树的构建
3. For node.childlist (i) in node
4. if (!e is a leaf-node)
5. return Insert (node.child(i), O_i)
6. Else
7. If (node.objects.length>M)//如果节点数量超过了 M 值
8. Split (node, newnode, O_i);} //调用分离函数
9. else
10. node.objects.Add (O_i); //加入到节点的对象集合中
11. update (node); //更新节点区域大小、关键字索引表
12. return Insert (node, O_i nextobject) //继续插入下一个节点
13. EndIf
14. EndIf

通过上述方法，我们构建了一个兼具语义特征和地理位置信息的高效资源索引结构，能够快速支持语义查询和优化查询性能。

2.2. 业务驱动的资源要素资源服务化匹配技术

智能家居产业链中的资源供需匹配通过云平台进行资源要素的搜索、发现与响应[7]。由于智能家居行业需求多样、信息量大且资源动态变化，匹配过程面临效率和精确度的挑战。为应对这些问题，提出了基于功能与性能相结合的资源匹配框架。该框架通过三个阶段：首先，通过功能型匹配筛选出满足基本需求的资源集；接着，通过性能匹配对资源进行优先级排序；最后，选出最优资源。此框架旨在提高资源匹配的效率和精确度，解决资源搜索与匹配中的复杂性[8]。

资源匹配流程基于业务驱动的供需框架，分为三个阶段，每个阶段在前一阶段的基础上细化匹配范围，避免全文搜索，从而提高匹配效率。第一阶段基于硬制造资源特征，如形状、数量、地理位置等，通过语义相似度算法进行初步筛选，得到初步服务集合。第二阶段针对软服务资源，通过匹配输入、输出和功能等属性，进一步筛选得到新的服务集合。第三阶段是 QoS 匹配，通过三角模糊数定量描述 QoS 指标，结合业务需求偏好，对服务集合进行排序，最终选出符合性能要求的资源。该过程通过逐步缩小匹配范围，确保了匹配的准确性和效率。

由于资源池中，资源要素服务化后基于语义信息进行聚类，因此计量语义相似度时主要依靠文本相似度，从词形、词序两个方面进行考虑。

A. 词形相似度

词形相似度可以反应两个文本形态上的相近程度，对于文本和文本，词形相似度计算方法如下：

$$\text{Sim}_w(T_1, T_2) = 2 * \frac{\text{SameWord}(T_1, T_2)}{\text{Len}(T_1) + \text{Len}(T_2)}$$

其中，SameWord 表示两个文本中相同单词的个数，Len 表示文本的长度，文本相同单词越多，表示两个文本词形相似度越高，Sim 值将越趋近于 1。

B. 词序相似度

词序相似度指两个文本中相同单词在文本中所处位置的相似程度，通过相邻分量逆序数来度量。对

于文本和文本，词序相似度计算方法如下：

$$\text{Sim}_o(T_1, T_2) = \begin{cases} 1 - \frac{\text{Revord}(T_1, T_2)}{|\text{OnceWord}(T_1, T_2)|}, & |\text{OnceWord}(T_1, T_2)| > 1 \\ \frac{1}{\max(\text{Len}(T_1), \text{Len}(T_2))}, & |\text{OnceWord}(T_1, T_2)| = 1 \\ 0, & |\text{OnceWord}(T_1, T_2)| = 0 \end{cases}$$

其中， $|\text{OnceWord}(T_1, T_2)|$ 表示两个文本中相同单词的个数，利用 $\text{P}_{\text{first}}(T_1, T_2)$ 表示 $\text{OnceWord}(T_1, T_2)$ 中的单词在文本 T_1 中的位置序号构成的一维向量， $\text{P}_{\text{second}}(T_1, T_2)$ 表示 $\text{P}_{\text{first}}(T_1, T_2)$ 中分量对应的单词在文本 T_2 中的位置序号构成的一维向量， $\text{Revord}(T_1, T_2)$ 表示 $\text{P}_{\text{second}}(T_1, T_2)$ 中相邻分量的逆序数。

2.3. 数据驱动的多维度用户画像技术

针对家居用户画像个性化、动态化、全面化的需求，综合运用用户日常交互操作、时间轨迹等个性化数据，研究数据驱动的多维度用户画像技术[9]。综合运用多维度用户画像生成的各类标签及相应权重信息作为决策因子，构建个性化的智能家居产品链服务平台。

数据驱动的多维度用户画像技术包括数据获取及预处理、画像标签抽取、数据画像处理、画像展示四个部分。以下重点阐述前三个部分。

2.3.1 数据获取及预处理

用户与智能家居平台进行的每一次交互行为中[10]，系统均会详细记录这些操作，并存储为结构化的操作日志。此外，家居环境中部署的摄像设备所捕获的图像视频资料以及红外传感器收集的数据，作为辅助信息，共同用于构建用户的详细时间线。在成功获取这些丰富的元数据之后，需依据预先设定的用户画像标签体系，对这些数据进行系统的归纳与整理。此过程涉及数据对齐与清洗等关键步骤，旨在剔除数据中的噪声成分，确保数据的准确性和一致性。最终，原始数据将被转化为一种标准化的格式，以便后续的数据处理与分析平台能够有效地接收并处理这些信息。

2.3.2. 画像标签抽取

对用户画像的主要目的是提升家居产品与用户交互的智能化水平，满足家居产品使用者的需求，同时要尽可能避免冗余。与独立的家居产品不同，智能家居产品链服务平台作为一个系统性整体，有着独特的画像标签，以供收集的数据能够在标签的基础上进行数据处理，生成用户画像[11]。

2.3.3. 数据画像处理

在数据对标签进行刻画处理的过程中，利用大数据处理架构 Hadoop 以及 Spark 框架，可以更加精确、高效地得到用户的具体个性化标签。利用 Hadoop 技术框架处理企业的海量数据，完成并行化的数据处理。利用全局熵值法进行具体计算。

具体的计算步骤如下：

① 对 t 天内的 n 个智能设备共 p 个指标进行评估，在数据收集及预处理后，得到 t 张数据表，每张数据表中为一个 $n \times p$ 格式的矩阵，之后将 t 张数据表按时间顺序排列，构成 $nt \times p$ 的全局矩阵，记作 $X_g = (X^1, X^2, X^3, \dots, X^t)$ 。

② 对全局矩阵 X_g 中的数据进行标准化，采用正向指标计算：

$$x'_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}, 1 \leq i \leq nt, 1 \leq j \leq p$$

当指标越大, 用户使用该设备的期望越强烈。

③ 计算第 j 个指标在第 i 个设备在该指标中所占比重: $f_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}}$ 。

④ 计算第 j 个指标的信息熵值: $e_j = -k \sum_{i=1}^m f_{ij} \ln f_{ij}$ 。

⑤ 计算第 j 个指标差异系数: $g_j = 1 - e_j$ 。

⑥ 计算各指标权重: $w_j = \frac{g_j}{\sum_{j=1}^p g_j}$ 。

⑦ 计算期望分值: $\sum_{j=1}^p w_j x'_{ij}$ 。

3. 平台创新以及优势

当前的数字化服务平台还存在着数字化产业链的局限性问题, 即平台应用服务较为局限, 无法对整条产业链的产品、数据和服务等进行协同管理, 如 Christou 等人的研究; 并且, 在数据接入到平台后, 无法对数据进行全链路的溯源、追踪和监控, 在数据的可视化分析和质量控制上仍存在问题, 如王昀等人、Liu 等人的研究。本文提出的智能家居产品链服务平台实现了全链条的智能生态服务, 提出了数据血缘的概念和数据全景监控功能, 有效解决了上述问题, 并创新性地引入了 AI 算法服务, 提供标准化的 AI 算法接口[11]。本文提出的智能家居产品链服务平台的创新及优势如下:

1) 全链条智能生态服务: 本研究构建了完整的智能家居生态链, 实现了从云存储到数据中台的全链条集成, 涵盖了硬件设备、数据管理和应用服务的智能化, 提供了硬件设备的数据收集、数据交互和数据监控功能和基于用户行为的个性化服务和智能决策支持, 打破了传统家居行业局限于单一设备或平台的局限, 使得跨平台、多设备的智能连接与管理成为可能。

2) 算法服务与 AI 集成: 为进一步提升智能家居平台的智能化水平, 本研究创新性地引入了 AI 算法服务模块, 提供标准化的 AI 算法接口, 支撑多维度的智能应用场景, 包括但不限于个性化服务、预警系统和用户行为分析等。通过算法服务模块的集成, 智能家居平台的使用体验得到了显著提升。用户可以享受到更加智能化的服务。

3) 数据血缘和数据全景监控: 利用数据血缘和全景监控功能, 该平台实现了从数据接入到应用的全链路追踪和监控。该功能解决了在传统家居行业中大规模数据的追踪和质量控制的难题, 为高效的数据管理和数据定位需求提供了技术支持。

本文提出的智能家居产品链服务平台集成了物联网、人工智能和云计算等数字化技术, 在相关研究的基础上, 扩展并改进了平台的应用服务功能, 解决了平台部署和产业链管理的局限性问题, 实现了数据流的全方位管理与全景监控, 并引入了 AI 算法对多维度场景的数据操作进行支持, 在多个场景下的测试均证明了其有效性。此外, 随着智能家居产业市场规模的不断扩大以及国家多项相关政策的支持, 智能家居产业链的数字化服务平台的建设发展也得到了保障。

4. 总结与展望

本研究针对智能家居产品链服务平台的数据中台架构进行了深入分析, 得出结论: 数据中台架构成功实现了数据源的整合与统一管理, 提高了数据处理的效率和准确性; 通过构建强大的数据管理和分析功能, 平台能够实时监控和优化智能家居设备的运行状态, 从而提升了用户体验和系统可靠性; 此外, 数据中台还为平台提供了灵活的数据访问和共享能力, 支持了个性化服务的创新和业务流程的优化。

尽管数据中台架构在提升平台性能和服务能力方面取得了显著成果, 但在实际应用中仍面临一些挑战, 包括数据隐私保护、系统扩展性以及多样化数据源的兼容问题。未来研究应重点关注如何进一步增

强数据安全性和隐私保护,提升数据中台对新兴技术的适应能力。同时,探索智能家居环境下的数据中台与人工智能、大数据技术的深度融合,将是实现更高效服务和智能化决策的关键方向。

基金项目

本研究由国家重点研发计划项目资助,项目编号 2021YFF0901200。

参考文献

- [1] 王一炜, 郑仁思, 宋嘉欣. 物联网趋势下的智能家居产品设计策略[J]. 轻纺工业与技术, 2021, 50(9): 103-104.
- [2] 王昀, 杜志双, 田康, 等. 数字政府时空数据中台设计与实践[J]. 大数据, 2024, 10(4): 149-160.
- [3] Liu, Z., Sampaio, P., Pishchulov, G., Mehandjiev, N., Cisneros-Cabrera, S., Schirrmann, A., *et al.* (2022) The Architectural Design and Implementation of a Digital Platform for Industry 4.0 SME Collaboration. *Computers in Industry*, **138**, Article 103623. <https://doi.org/10.1016/j.compind.2022.103623>
- [4] Christou, I.T., Kefalakis, N., Soldatos, J.K. and Despotopoulou, A. (2022) End-to-End Industrial IoT Platform for Quality 4.0 Applications. *Computers in Industry*, **137**, Article 103591. <https://doi.org/10.1016/j.compind.2021.103591>
- [5] Mao, Z., Wu, J., Qiao, Y. and Yao, H. (2021) Government Data Governance Framework Based on a Data Middle Platform. *Aslib Journal of Information Management*, **74**, 289-310. <https://doi.org/10.1108/ajim-03-2021-0068>
- [6] Pamucar, D., Simic, V., Görçün, Ö.F. and Küçükönder, H. (2024) Selection of the Best Big Data Platform Using COBRAC-ARTASI Methodology with Adaptive Standardized Intervals. *Expert Systems with Applications*, **239**, Article 122312. <https://doi.org/10.1016/j.eswa.2023.122312>
- [7] 吴梦莹. 工业互联网中台管理系统的架构设计与实现[J]. 电子通信与计算机科学, 2024, 6(7): 178-180.
- [8] Guo, T., Han, J.T. and Yang, X.L. (2015) Architecture and Key Capabilities of Management System of Smart Home Device. *Telecommunications Science*, **31**, 110-115.
- [9] 严萍, 张兴敢, 柏业超, 杜仲林. 基于物联网技术的智能家居系统[J]. 南京大学学报(自然科学版), 2012, 48(1): 26-32.
- [10] 何文丰. 一种多层结构型智能家居系统的设计与实现[D]: [硕士学位论文]. 厦门: 厦门大学, 2023.
- [11] 朱峰. 基于智能家居环境下信息监测系统的设计开发[J]. 电子通信与计算机科学, 2022, 4(3): 121-122.