# 软件定义物联网中多源异构数据智能融合处理研究

魏枞浩

（北京邮电大学，北京 100876）

**摘　要：** 对于多源异构数据融合，现有方法通常依赖单一特征进行匹配，难以充分利用不同数据源之间的关联关系，导致融合结果存在偏差。文章设计了一种面向软件定义物联网的多源异构数据智能融合方法，首先建立多源异构数据语义映射关系，通过语义对齐降低不同数据源之间的表达差异；其次基于软件定义架构设计对齐数据智能融合算法，利用权重分配实现多源数据初步融合；最后引入并行处理机制，对同区域内多个关联数据目标进行协同计算，从而获得综合融合结果。实验结果表明，采用该方法进行数据融合时，LSE值均低于0.3，较对比方法具有更小的融合误差，能够有效提高多源异构数据融合精度。

**关键词：** 软件定义；物联网；多源异构数据；数据融合；分类处理；数据源

**中图分类号：** TP311　　**文献标识码：** A

# Research on intelligent fusion processing of multi-source heterogeneous data in software-defined Internet of Things

WEI Zonghao

（Beijing University of Posts and Telecommunications，Beijing 100876，China）

**Abstract**: Existing multi-source heterogeneous data fusion methods generally rely on single features for matching, making it difficult to fully exploit the correlations among different data sources and resulting in biased fusion results. This paper designs an intelligent fusion method for multi-source heterogeneous data in software-defined Internet of Things. First, semantic mapping relationships among multi-source heterogeneous data are established, and semantic alignment is performed to reduce expression differences among different data sources. Second, an intelligent fusion algorithm for aligned data is designed based on the software-defined architecture, where weight allocation is applied to achieve preliminary multi-source data fusion. Finally, a parallel processing mechanism is introduced to perform collaborative computation on multiple associated data targets within the same region, thereby obtaining comprehensive fusion results. Experimental results show that the LSE values obtained by the proposed method are all below 0.3, indicating smaller fusion errors compared with the comparison methods. The proposed method can effectively improve the accuracy of multi-source heterogeneous data fusion.

**Key words**: software-defined, Internet of Things, multi-source heterogeneous data, data fusion, classification processing, data source

## 1　引言

随着物联网设备数量不断增加和应用场景持续扩展，物联网系统产生的数据类型和数据来源日益多样化。然而，不同来源数据在结构、格式和语义等方面存在明显差异，使得多源异构数据融合成为物联网数据处理中的重要问题。现有融合方法通常依赖特征提取与匹配实现数据整合，在处理多目标关联数据时，难以充分利用不同数据源之间的关联关系，导致融合结果存在一定的偏差。

王战华等[1]采用深度神经决策森林方法，通过主元分析法进行降维并提取特征实现融合。该方法融合效率较高，但多目标融合过程中仍存在关联信息利用不足的问题。郭鹏等[2]设计了多源异构网络安全数据可视化融合方法，利用Apriori算法进行关联挖掘，提高了网络安全数据分析效率，但该方法尚未统一不同类型数据之间的关联标准。李莉等[3]面向电力气象数据，采用改进SVM与WOA算法提取特征，提高了融合过程的稳定性，但特征引导方式在复杂数据场景下存在一定局限。张凌志[4]基于联合卡尔曼滤波，通过插值修复与层级融合处理多批次数据。该方法能够同步处理不同批次数据，但在层级修复过程中可能引入额外误差。

软件定义物联网通过软件化控制方式提高了设备管理和数据处理的灵活性，为多源异构数据融合提供了

统一处理环境。基于此，本文设计了一种面向软件定义物联网的多源异构数据智能融合方法。

## 2 软件定义物联网多源异构数据智能融合方法设计

### 2.1 多源异构数据语义映射与对齐

为降低不同数据源之间的表达差异，在进行数据融合前，需要对异构数据进行语义关联分析。对此，建立面向物联网的异构数据集，通过语义向量计算数据间相似程度，并构建语义映射关系[5-6]：

$$F_B=B\times\frac{1}{Q}-\frac{|\overrightarrow{\boldsymbol{v}_{c1}}\cdot\overrightarrow{\boldsymbol{v}_{c2}}|}{|\boldsymbol{v}_{c1}||\boldsymbol{v}_{c2}|} \tag{1}$$

$$\begin{cases} K=(v_{c1}+v_{c2})-F_B\iota\left(B\times\dfrac{1}{Q}\right) \\ K'=\iota Q\times\delta+\dfrac{\boldsymbol{v}_{c1}+\boldsymbol{v}_{c2}}{F_B} \end{cases} \tag{2}$$

其中，$F_B$表示语义相似度；$B$表示语义权重；$Q$表示关联数据；$\boldsymbol{v}_{c1}$，$\boldsymbol{v}_{c2}$分别表示前后关联向量；$|\boldsymbol{v}_{c1}|$，$|\boldsymbol{v}_{c2}|$分别表示对应向量的模；$K$，$K'$分别表示异构数据映射规则；$\iota$表示相似目标；$\delta$表示相似度阈值。

基于建立的语义映射规则，对多源异构数据进行语义对齐处理[7]。对于存在映射关系的数据，将其语义统一为目标数据源中的表达形式[8]。实际应用过程中，不同数据源可能对同一概念存在不同定义，导致语义冲突。针对该问题，采用加权投票方法，根据不同数据源权重确定最终语义结果，实现冲突数据对齐：

$$c^Q=\operatorname{argmax}_{c_i}\sum_{j=1}^{n}w_j\cdot I-\frac{(\boldsymbol{v}_{c1}+\boldsymbol{v}_{c2})}{K'} \tag{3}$$

其中，$c^Q$表示对齐后语义；$c_i$表示存在的冲突目标；$n$表示最大对齐目标；$j$表示目标权重；$w_j$表示非冲突数据；$I(\cdot)$表示指示函数。

### 2.2 基于软件定义架构的数据融合算法设计

基于软件定义技术，设计面向物联网异构数据的数据融合算法。由于物联网数据处于持续更新状态，在完成数据对齐处理后，需要对输入数据进行预处理，以保证后续融合过程的数据质量。数据经过RCMS、路由器和交换机等设备完成接入与传输，随后由RCMS进行统一处理（包括格式标准化、缺失值处理和异常值检测等），从而为后续数据融合提供有效输入。

RCMS预处理后的数据采用熵权法进行权重计算。首先，确定不同数据源的信息熵，并根据熵值计算数据源权重：

$$E_j=-\frac{1}{\ln n}\sum_{j=1}^{n}p_{ij}\cdot\ln p_{ij}-c^Q(\boldsymbol{v}_{c1}+\boldsymbol{v}_{c2}) \tag{4}$$

$$P=\frac{E_j}{|\boldsymbol{v}_{c1}||\boldsymbol{v}_{c2}|}\cdot\ln p_{ij}\left(w_j\right)^I \tag{5}$$

其中，$E_j$表示数据源熵值；$p_{ij}$表示实时数据占比；$P$表示权重分配；$p_{ij}\left(w_j\right)^I$表示临近分配目标。其次，采用加权平均方式对各数据源的对齐数据进行融合计算。在软件定义架构下，根据熵权法确定不同数据源在融合过程中的贡献程度；利用计算得到的权重对完成语义对齐的数据进行加权处理，从而输出初步融合结果。

### 2.3 基于并行处理的数据融合优化

并行处理通过对多个周期内的融合目标进行协同计算，提高多源异构数据处理效率。由于多目标、多数据处理过程中存在较大的计算负载，需要进行负载均衡处理，其计算过程如下：

$$L_P=\alpha\cdot t_P+\gamma-\frac{K'\cdot B\cdot\ln p_{ij}\left(w_j\right)^I}{|\boldsymbol{v}_{c1}||\boldsymbol{v}_{c2}|} \tag{6}$$

其中，$L_P$表示负载均衡；$\alpha$表示负载指数；$t_P$表示汇总融合目标；$\gamma$表示更新目标数量。

完成动态负载调整后，接入主从式并行架构。其中，主节点负责整体任务调度、数据分配和结果汇总，从节点负责执行具体的数据处理与融合计算。基于该架构，各从节点根据分配的数据集和计算得到的权重，并行完成多目标融合，计算过程如下：

$$Y_{t_p}=\frac{1}{n}\times\ln p_{ij}\left(w_j\right)^I-\frac{(\boldsymbol{v}_{c1}+\boldsymbol{v}_{c2})}{L_P\cdot\upsilon} \tag{7}$$

其中，$Y_{t_p}$表示并行融合结果；$\upsilon$表示全局融合目标数量。并行处理可有效减轻单节点数据处理压力，提高多源异构数据融合效率，同时减少融合过程中的计算偏差。

## 3 实验分析

### 3.1 实验环境与参数设置

为验证本文方法的有效性，基于L物联网平台开展测试实验。实验平台接入Ryu或ONOS开源控制器，并部署于Ubuntu服务环境中，模拟软件定义物联网（SD-IoT）运行环境。平台通过Open vSwitch虚拟交换机和REST API实现异构数据采集与状态监测，并部署数据监测模块对存储数据和更新数据进行采集。该模块通过REST API与Open vSwitch虚拟交换机交互，以获取数据规模、更新频率和来源设备等状态信息，为后续数据融合提供数据依据。

实验过程中，采集不同类型物联网数据作为输入数据集，具体数据规模如表1所列。

**表1 物联网异构数据集**

| 名称 | 数据规模/条 |
|---|---|
| Env-Sensor | 23 000 |
| Device-Log | 20 000 |
| User-Behavior | 2 470 |
| Image-Surv | 13 000 |
| Power-Grid | 2 835 |
| Health-Wearable | 1 200 |

将表 1 所列数据作为异构数据源，在软件定义环境下开展融合性能测试。为保证实验条件一致，设置统一的数据融合参数，具体如表 2 所列。

表 2　异构数据实践融合参数

| 实践指标 | 参数标准 |
|---|---|
| 权重分配 | 5.20 |
| 数据归一化范围 | [0,1] |
| 学习率 | 0.01 |
| 采样频次/h | 4 |

### 3.2　数据关联性分析

选取 4 批次异构数据，计算各批次数据之间的相关系数，并对数据关联特征进行分析。分析结果表明，第一批次数据的相关系数处于 0.4~0.8 范围内，说明不同数据源之间具有一定关联性；第二批次与第三批次数据的相关系数均高于 0.7，整体变化较为稳定，表明两批次数据具有较高关联程度；第四批次数据的相关系数处于 0.3~0.6 范围内，整体低于前三批次，且部分数据关联系数低于 0.5，说明该批次中存在关联程度较弱的数据，需要进一步筛选处理。

### 3.3　融合性能对比分析

基于实验数据集，采用本文方法、深度神经决策森林信息化异构多元目标数据融合方法以及多源异构网络安全数据可视化融合方法开展对比实验，结果如图 1 所示。

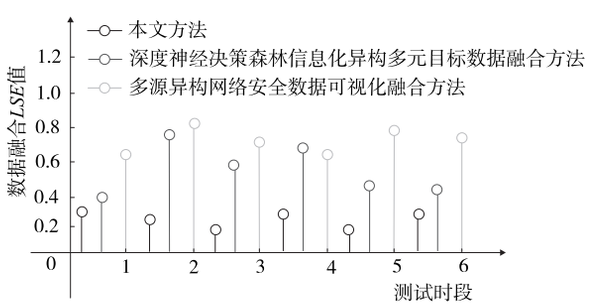


图 1　数据融合 LSE 值结果对比

实验结果表明，采用本文方法进行数据融合时，*LSE* 值均低于 0.3，而两种对比方法的 *LSE* 值分别处于 0.4~0.8 和 0.5~0.9 范围内。*LSE* 值越小表示融合误差越低，说明采用本文方法能够获得更小的数据融合误差。该实验结果表明，语义映射能够减少不同数据源之间的表达差异，权重融合能够提高关联数据的利用程度，并行处理能够增强多目标数据的协同计算能力，三者共同作用减小了数据融合过程中的误差。

## 4　结束语

本文围绕软件定义物联网中的多源异构数据融合问题，设计了基于语义映射、权重融合和并行处理的数据融合方法。通过建立数据间语义关联关系，减小不同来源数据之间的表达差异，并结合并行计算提高融合处理能力。实验结果表明，采用本文方法进行数据融合时，*LSE* 值均低于 0.3，有效减小了融合误差，提高了融合结果准确性。后续将面向动态物联网环境，进一步强化该方法的数据实时处理和多设备协同融合能力。

**参考文献：**

[1] 王战华，黄喜悦．基于深度神经决策森林的信息化异构多元目标数据融合方法[J]．微型电脑应用，2025，41(6)：285-287.

[2] 郭鹏，袁飞．多源异构网络安全数据可视化融合分析方法[J]．现代计算机，2024，30(22)：146-150.

[3] 李莉，孙世军，朱坤双，等．面向电力气象数据的多源异构数据融合方法研究[J]．电子设计工程，2024，32(16)：178-182.

[4] 张凌志．基于联合卡尔曼滤波的异构网络数据信息融合方法[J]．长江信息通信，2024，37(5)：13-15.

[5] 段昕汝，陈桂茸，姬伟峰，等．基于联邦学习的多源异构网络无数据融合方法[J]．空军工程大学学报，2024，25(1)：90-97.

[6] 谭如超，何群，王华，等．异构网络通信数据融合方法研究[J]．自动化仪表，2023，44(7)：50-54+60.

[7] 陈征，高明亮，蒋卫国，等．面向污染场地智能化管控多源异构数据融合方法综述[J]．黑龙江科学，2023，14(8)：1-6+12.

[8] 张红蕾．基于物联网技术的网络异构传感数据融合方法[J]．信息与电脑(理论版)，2023，35(3)：201-203.

**作者简介：**

魏枞浩(2005—)，本科，研究方向：软件工程。