引文格式:李琳慧,王宇,刘越岩,等.特征相似度驱动的不动产多源异构数据快速融合模型[J].武汉大学学报(信息科学版),2025,50(7):1447-1454.DOI:10.13203/j.whugis20220742

Citation: LI Linhui, WANG Yu, LIU Yueyan, et al. A Fast Fusion Model for Multi-source Heterogeneous Real Estate Data Driven by Feature Similarity[J]. Geomatics and Information Science of Wuhan University, 2025, 50(7):1447-1454.DOI:10.13203/j.whugis20220742

特征相似度驱动的不动产多源异构数据快速融合模型

李琳慧[1] 王 宇[2] 刘越岩[2] 李 磊[1] 黄锦丞[2] 周 熠[2] 曹淞霖[2]

1 武汉市房产测绘中心,湖北 武汉,430015

2 中国地质大学(武汉)公共管理学院,湖北 武汉,430074

摘 要:多源异构数据快速融合是不动产统一登记过程中数据整合与建库面临的难点问题。现有的不动产多源异构数据融合方法无法准确构建空间实体间的关联关系,且对数据规模庞大的不动产数据存在时间成本过高的问题。通过引入特征相似度技术,提出了一种多层级快速融合模型,实现不动产多源异构数据的批量融合。首先,在对不动产多源异构数据进行电子化和矢量化的基础上,引入相似度因子对多源异构数据进行评估;然后,设计了一种基于综合相似度加权算法的网络模型,计算不动产多源异构数据在各相似度因子分布方向上的综合相似度;最后,采用相似度阈值参数和限定范围参数进一步提高多源异构数据融合的精度和效率。以不动产数据为例,对快速融合模型进行定量化分析,实验结果表明,相比于其他方法,所提模型时间成本低,每千幢的平均融合时间约为1 s,融合准确率达到93.50%,能有效提升不动产多源异构数据的融合精度与效率。

关键词:多源异构数据;特征相似度;快速融合;相似度因子;多层级;不动产

中图分类号:P208;P273 **文献标识码**:A **收稿日期**:2023-03-19

DOI:10.13203/j.whugis20220742 **文章编号**:1671-8860(2025)07-1447-08

A Fast Fusion Model for Multi-source Heterogeneous Real Estate Data Driven by Feature Similarity

LI Linhui[1] *WANG Yu*[2] *LIU Yueyan*[2] *LI Lei*[1] *HUANG Jincheng*[2]
ZHOU Yi[2] *CAO Songlin*[2]

1 Wuhan Real Estate Surveying and Mapping Center, Wuhan 430015, China

2 School of Public Administration, China University of Geosciences(Wuhan), Wuhan 430074, China

Abstract: **Objectives**: Fast fusion of multi-source heterogeneous data of real estate is a difficult problem faced by data integration and database construction in the process of unified real estate registration. The existing multi-source heterogeneous data fusion methods of real estate are unable to accurately construct the relationship describing the spatial entities, and the time cost is too high for the real estate data with huge data scale. **Methods**: By introducing feature similarity technology, we propose a multi-level fast fusion model to realize batch fusion of real estate multi-source heterogeneous data. First, on the basis of electronization and vectorization of real estate multi-source heterogeneous data, similarity factor is introduced to evaluate real estate multi-source heterogeneous data. Then, a network model based on comprehensive similarity weighting algorithm is designed to calculate the comprehensive similarity of real estate multi-source heterogeneous data in the distribution direction of each similarity factor. Finally, similarity threshold parameters and limited range parameters are used to further improve the accuracy and efficiency of multi-source heterogeneous data fusion. **Results and Conclusions**: We take real estate data as an example to conduct quantita-

第一作者:李琳慧,硕士,主要从事地图制图和地理信息系统研究。499535459@qq.com

通信作者:刘越岩,硕士,副教授。yueyanliu@cug.edu.cn

tive analysis of the fast fusion model. The experimental results show that compared with other methods, the proposed model consumes less time and costs, the average fusion time per thousand houses is 1 s, and the fusion accuracy reaches 93.50%, which can effectively improve the fusion accuracy and efficiency of multi-source heterogeneous data of real estate.

Key words: multi-source heterogeneous data; feature similarity; fast fusion; similarity factor; multiple level; real estate

在不动产统一登记的过程中,以不动产数据为典型代表的多源异构数据存在数据种类繁多、数据规模庞大、数据信息不全、数据格式不一致[1]等问题,且各数据系统间缺乏有效关联,数据信息的孤立制约着国家大数据发展战略的进步。多源异构数据融合是解决数据孤岛和数据集成的有效途径,并逐渐成为数据科学领域的研究热点。

多源异构数据融合技术在目标跟踪[2]、图像处理[3]、地质勘探[4]、健康医疗[5]、信息系统[6]、城市感知[7]等众多科学领域中有着广泛的应用。各领域研究者在多源异构数据融合方面开展了大量研究工作和有益尝试。传统多源异构数据融合算法多以机器学习算法为主,如基于改进粒子群优化的算法[8]、联邦学习框架算法[9]、联合卡尔曼滤波算法[10]以及其他主流机器学习算法[11],机器学习算法具备较高的精度,但对于数据规模庞大的多源异构数据,其运算效率较低。为提升多源异构数据融合的效率,基于图神经网络的算法[12]、改进残差网络算法[13]等神经网络算法不断涌现。神经网络算法具备较高的精度与效率,但前期训练成本较高,且对不同类型数据的迁移学习效率较差。同时,其他算法如深度泛化模型[14]、数据同化算法[15]以及其他多源异构数据融合算法[16-18]也拥有较高的精度,但仅适用于特定数据集或小型数据集。

现有的不动产多源异构数据融合方法多面向非空间数据融合情景,是对属性数据的融合,多基于机器学习算法和神经网络算法。然而,针对空间数据和非空间数据间融合的研究方法较少,且多基于可扩展标记语言(extensible markup language,XML)[19-20]和特征操作引擎(feature manipulate engineering,FME)[21-22]等方法。由于各不动产数据库建库时间点不一致,部分数据库未及时更新数据信息,不动产非空间数据不具备数据现势性,且不动产空间数据多存在平移、旋转、缩放等问题,融合精度差强人意,对于数据规模庞大、数据类型复杂[23]的不动产数据,需耗费高额时间成本,因此,上述方法并非完全适用。如何精准构建关联关系是提升不动产空间数据与非空间数据间融合精度和效率的关键。

为增强空间数据和非空间数据融合情景下的融合精度和效率,针对现有多源异构数据融合方法的不足,本文提出了一种基于特征相似度的不动产多源异构数据快速融合模型,并以不动产数据为例,实现不动产空间数据与非空间数据融合情景下精准关联关系的构建与不动产数据的批量融合,提高不动产数据融合精度与效率,为多源异构数据融合提供一种高效的实现方法。

1 数据与方法

1.1 实验数据

不动产数据拥有数据规模庞大、数据标准不一致、数据种类繁多等特点,同时具备空间数据与非空间数据,是多源异构数据的典型代表。本文以不动产数据为多源异构数据样本,其核心分为以下两部分:

1)权籍数据库,包含权利人信息、房屋坐落、房屋面积等信息,该数据主要以非空间数据形式存储。部分数据生成时期较早,采用纸质形式存放,数据存在房屋坐落信息缺失或不全情况,且由于街道拆并、地名变更、房屋拆迁、门牌号重编以及数据未及时更新等因素,部分数据或存失真可能。

2)房产管理空间数据库,包含房屋平面图、宗地图等信息,该数据主要以空间数据形式存储。该数据库建库时间点较新,拥有最新且完整的房屋数据信息,但数据库建立时期并未采集相应权利人信息,无法确定房屋权属性质。

1.2 多源异构数据融合流程

1.2.1 特征数据提取

特征数据分为空间特征数据和非空间特征数据,空间特征数据以房产管理空间数据库为数据源,采用栅格转矢量技术提取空间平面图、宗地图和空间坐标等空间数据信息构建相应特征数据库;非空间特征数据以权籍数据库为数据源,采用光学字符识别(optical character recognition,OCR)等技术提取房屋权利人、房屋坐落、房

屋面积、房屋楼层等非空间数据信息构建相应特征数据库。对于部分生成时期较早的纸质数据,采用电子扫描等技术,在电子化的基础上再进行特征数据提取及相应特征数据库构建。如在电子化纸质分层分户图的基础上,提取房屋坐落信息并构建地名地址数据库等。

1.2.2　快速融合模型构建

快速融合模型能精准构建不动产空间数据与非空间数据间的关联关系,通过关联关系对权籍数据库和房产管理空间数据库进行数据融合,形成统一的不动产数据。同时,快速融合模型能降低关联过程和融合过程中的时间成本,实现不动产多源异构数据的快速融合。

依据空间与非空间特征数据库中的特征数据选取合适相似度因子,为各相似度因子分配权重系数,按照相似度因子运算最优序列构建相似度因子筛选层,并计算多源异构数据在各相似度因子分布方向上的综合相似度,对综合相似度降序排列并生成融合结果列表输出。

相似度因子是描述目标对象之间稳定且一致特征数据的参数,是构建快速融合模型的基础。它用于衡量空间与非空间数据集之间的相似程度。依据多源异构数据的不同特性和实际需要,可选取不同相似度因子构建快速融合模型。

相似度因子选取需遵循以下选取原则:选取个数少,计算难度低,数据特征代表程度高,具备平移、旋转、缩放不变性。根据上述选取原则,本文针对不动产数据的诸多特性,考虑人力、时间成本,从空间与非空间特征数据库中分别选取两个特征数据作为相似度因子,空间特征数据库包括形状指数(shape index,SI)和狭长度,非空间特征数据库包括楼层数和面积。本文共选取4个特征数据作为相似度因子。部分相似度因子计算原理如下:

1)多边形面积相似度(square similarity,SS)因子。设构成 m 边形数据的顶点坐标集为 $P=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$,其中,$(x_m,y_m)$ 为多边形第 m 个顶点的坐标,则多边形面积相似度的计算公式为:

$$S_{SS}=\frac{A_2}{A_1},0\leqslant S_{SS}\leqslant 1 \tag{1}$$

$$A=\sum_{j=1}^{m}\frac{1}{2}(x_j y_{j+1}-y_j x_{j+1}) \tag{2}$$

式中,S_{SS} 为多边形面积相似度;A 为多边形面积,A_1、A_2 分别为权籍数据库和房产管理空间数据库中的多边形面积;(x_j,y_j) 为多边形第 i 个顶点的坐标;m 为多边形顶点数量。

2)多边形形状指数相似度(shape index similarity,SSI)因子。形状指数是指计算区域内某图斑的实际形状与标准形状(正方形或者圆)之间的偏离程度[24]。

本文以正方形为标准形状,多边形形状指数相似度的计算公式为:

$$S_{SSI}=\frac{I_2}{I_1} \tag{3}$$

$$I=0.25\times C\times S^{-\frac{1}{2}} \tag{4}$$

$$C=\sum_{i=1}^{m}\sqrt{(x_i-x_{i+1})^2+(y_i-y_{i+1})^2} \tag{5}$$

式中,S_{SSI} 为多边形的形状指数相似度;I 为形状指数,I_1、I_2 分别为权籍数据库和房产管理空间数据库中多边形的形状指数;C 为多边形周长。I 越接近1,则表明多边形与标准形状的偏离程度越小;反之,则形状越复杂。

3)多边形狭长度相似度(aspect ratio similarity,SAR)因子。多边形狭长度是指某多边形最小外接矩形的长宽比。多边形狭长度相似度的计算公式为:

$$S_{SAR}=\frac{R_2}{R_1} \tag{6}$$

$$R=L\times W^{-1} \tag{7}$$

式中,S_{SAR} 为狭长度相似度;R 为多边形狭长度,R_1、R_2 分别为权籍数据库和房产管理空间数据库中多边形的形状指数;L、W 分别为最小外接矩形的长度、宽度。

选取相似度因子构建快速融合模型,快速融合模型的综合相似度(integrated similarity,IS)计算公式为:

$$S_{IS}=\sum_{i=1}^{n}w_i\times S_i \tag{8}$$

式中,S_{IS} 为模型融合的综合相似度;S_i 为模型中第 i 个相似度因子的相似度;w_i 为第 i 个相似度因子的权重系数;n 为相似度因子的个数。

1.2.3　模型参数调整

多源异构数据存在数据规模庞大、数据类型复杂、数据标准不一致的特性,且区域之间数据差异较明显。为提升模型融合的准确性,并使模型适用于更多种类的多源异构数据,需调整相似度因子的权重系数、相似度阈值参数和限定范围参数,从而优化模型。

相似度因子的权重系数采用层次分析法进行计算分配,由于区域间数据的差异性,权重系

数会因区域的不同而发生变化。本文通过对某区域内的样本数据进行实验来确定各相似度因子的权重系数。

快速融合模型根据数据的差异程度将不动产数据划分为5个级别,通过对某区域内的样本数据进行实验,以精度和效率的综合最优级别为该模型的当前最优相似度阈值。模型经过样本实验,以0.9的相似度阈值为模型的当前最优相似度阈值。

限定范围参数在保证融合精度的前提下能有效减少模型运算量,提高模型的融合效率。本文在地名地址数据库中,将坐落信息按照市、区、乡(镇)、街、社区、楼幢数、单元号、门牌号的分级标准进行数据整理。通过地名地址数据库提取相应的房屋坐落信息,以该房屋坐落信息为基准确定模型最小限定范围。如地名地址数据库中某房屋坐落信息不全,仅具体到庙山街道,则限定范围以庙山街道内所有的不动产数量为模型最小限定范围。

1.2.4 多源异构数据融合

融合结果列表仅有唯一备选项,人工确认无误后对空间与非空间数据进行融合;融合结果列表存在多项备选项,人工剔除错误选项后,对空间与非空间数据进行融合。若无正确备选项,则表示部分数据存在缺失不全或失真,需人工查阅档案并外业采集核实不动产信息后,将数据录入数据库完成信息更新。多源异构数据快速融合模型的主要流程如图1所示。

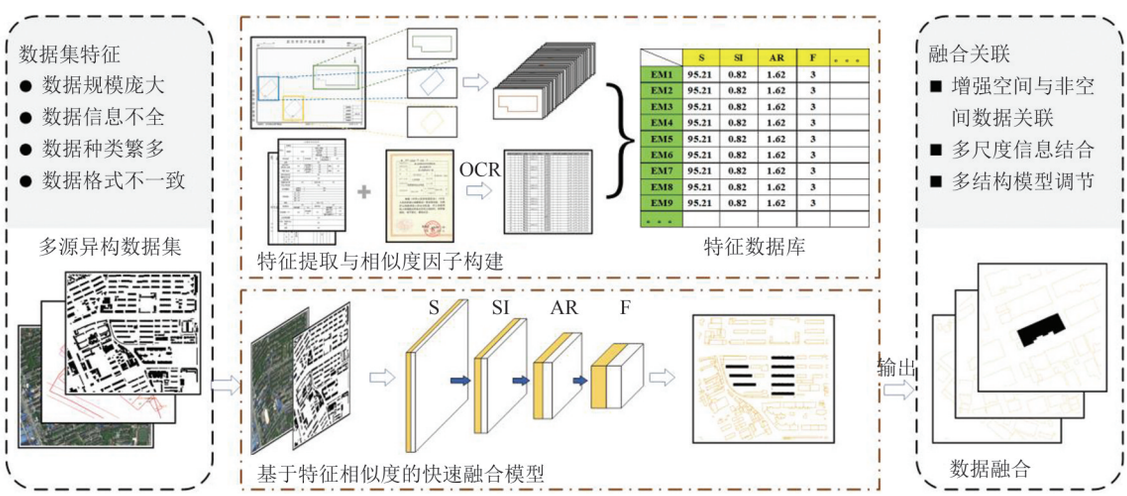


图1 多源异构数据快速融合模型的主要流程示意图

Fig. 1 Diagram of Main Flow of Fast Fusion Model of Multi-source Heterogeneous Data

2 不动产数据融合结果与评价

2.1 评价方法与指标

为验证快速融合模型的性能,本文从精度和效率两方面对模型进行评估,并与主流的多源异构融合算法进行对比。

从权籍数据库中共选取10 000条样本数据,按照1:1的比例分为正向样本数据集和负向样本数据集,并采用准确率A、精度P、召回率R和综合评价指标F来评估模型的性能。

准确率A是基于混淆矩阵的衍生评价指标,被分类为正确的样本数量占全体样本数量的比值,其计算公式为:

$$A=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{FN}+N_{FP}+N_{TN}}\times 100\% \quad (9)$$

式中,TP(true positive)为真正样本,N_{TP}为真实值中所有正向样本分类正确的样本数量;TN(true negative)真负样本,N_{TN}为真实值中所有正向样本分类错误的样本数量;FP(false positive)为假正样本,N_{FP}为真实值中所有负向样本分类错误的样本数量;FN(false negative)为假负样本,N_{FN}为真实值中所有负向样本分类正确的样本数量。

精度P是指被分类为正向样本的样本数量占正向样本总数的比值,以二分类混淆矩阵为例,其计算公式为:

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}\times 100\% \quad (10)$$

综合评价指标F是精度P和召回率R的加权调和平均,能更为全面地评价模型分类的能力。其计算公式为:

$$F=\frac{(\alpha^2+1)\times P\times R}{\alpha^2\times(P+R)}\times 100\% \quad (11)$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}}\times 100\% \tag{12}$$

式中,α是调和参数;R为召回率。

2.2　算法精度分析

为检验快速融合模型的性能,本文共挑选16组样本数据,在8组不同限定范围的情景下,通过快速融合模型在样本数据与融合数据间构建精准的关联关系并进行数据融合。快速融合模型的结果如表1所示。从表1可知,实验结果表现良好,16组样本数据与房产管理空间数据库间全部构建出关联关系,且通过正向样本数据集真阳性验证,其个别样本数据出现非唯一性关联。

选取相同样本数据集,将本文模型与可扩展标记语言(extensible markup language,XML)、特征操作引擎(feature manipulate engine,FME)、快速特征检测和描述算法(oriented fast and rotated brief ,ORB)、平方差匹配法(square difference matching,TM_SQDIFF)、相关匹配法(correlation matching,TM_CCORR)、系数匹配法(correlation coefficient matching, TM_CCOEFF) 6种算法在多源异构数据融合的性能进行对比,结果见表2与图2。评估结果表明,本文模型在空间数据与非空间数据融合情景下取得较好的结果;其他6种算法在单一空间数据或非空间数据情景下虽然能取得较为优异的结果,但在空间数据与非空间数据融合情景下,相比本文模型,无法有效利用更多特征数据,算法性能表现较差,且模板匹配等算法对具备旋转、缩放特性的多源异构数据并不适用。

2.3　算法效率分析

为检验快速融合模型的效率,本文共挑选16组样本数据,在8组不同数据量的情景下,通过快速融合模型对其运行时间进行分析评价,快速融合模型的融合效率如表3所示。可以看出,实验结果整体良好,模型在实验中每千幢的平均融合时间为1 s,相似度整体维持在80%左右。

设置不同数据量,本文模型与其他6种算法在多源异构数据融合效率的性能对比结果如图3所示。由图3可知,基于XML、FME的算法及模板匹配算法具有较高的效率,但它们并不适用于空间数据与非空间数据情景下的多源异构数据融合,精度结果表现较低。相比之下,本文模型兼顾精度与效率,更加适用于该情景下的多源异构数据融合。

为挖掘多源异构数据融合流程的内部相关性,本文将多源异构数据融合流程划分为3个阶段,分别为特征数据提取阶段、关联关系构建阶段和数据融合阶段,如图4(a)所示。各多源异构数据融合方法拥有各自适用的领域,以不动产数据为例,现在的不动产空间数据和非空间数据融合多采用人工融合的方法,将本文模型与人工融合方法的效率进行对比,结果如图4(b)所示。

表1　快速融合模型精度结果

Table 1　Precision Results of Fast Fusion Model

实验数据	样本数据	房产管理空间数据库	限定范围/幢	选取数量/幢	是否构建关联
1			1 225	1	是
2			1 225	1	是
3			890	1	是
4			890	1	是
5			3 695	3	是
6			3 695	1	是
7			1 802	1	是
8			1 802	1	是
9			1 356	1	是
10			1 356	1	是
11			1 765	1	是
12			1 765	1	是
13			52 881	1	是
14			52 881	6	是
15			5 672	1	是
16			5 672	1	是

通过分析各个阶段信息,检验本文模型及人工融合方法的性能并应用于实际工作,各阶段结果如图5所示。实验以多次对比的均值为最终结果,在特征数据提取阶段,相比于人工融合方法,

表2 主流算法性能对比表/%

Table 2 Performance Comparison of Mainstream Algorithms/%

算法	准确率	精度	召回率	综合评价指标
XML	70.50	59.20	76.48	66.74
FME	74.90	62.40	83.20	71.31
ORB	85.10	77.20	91.68	83.82
TM_SQDIFF	54.80	37.20	57.41	45.15
TM_CCORR	57.00	43.60	59.83	48.77
TM_CCOEFF	65.40	49.60	72.51	58.91
本文模型	93.50	90.80	95.98	93.32

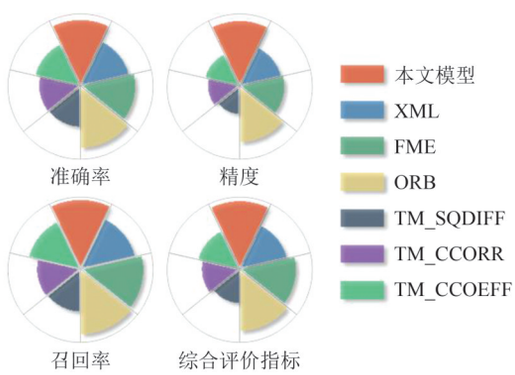


图2 主流算法性能对比图

Fig. 2 Diagram of Performance Comparison of Mainstream Algorithms

表3 快速融合模型效率结果

Table 3 Efficiency Results of Fast Fusion Model

实验数据	选取数/幢	限定范围/幢	融合时间/s	相似度/%
1	1	1 225	6.63	80.68
2	1	1 225	6.57	79.36
3	1	890	4.68	85.69
4	1	890	5.26	84.32
5	3	3 695	6.80	84.52
6	1	3 695	8.17	82.69
7	1	1 802	6.23	78.69
8	1	1 802	5.17	81.57
9	1	1 356	6.49	80.14
10	1	1 356	6.36	79.27
11	1	1 765	5.78	78.65
12	1	1 765	6.09	85.63
13	1	52 881	22.30	86.35
14	6	52 881	22.56	79.61
15	1	5 672	9.78	81.53
16	1	5 672	9.35	82.62

本文模型耗费更多的时间,这是由于在本文模型中,电子化、矢量化纸质数据和提取特征数据等数据处理过程耗费较多时间,而人工融合方法降低了数据处理过程的时间,使二者之间产生了一定的时间成本差额。在关联关系构建阶段,人工融合方法所耗时间明显大于模型匹配所耗时间,且二者之间的时间成本差额会随数据量的增大呈指数形式扩大。在数据融合阶段,二者所耗时间相差不大。

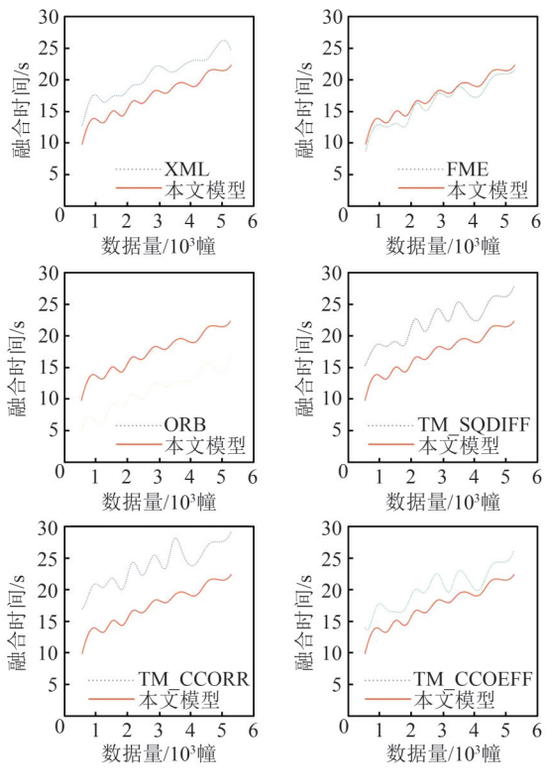
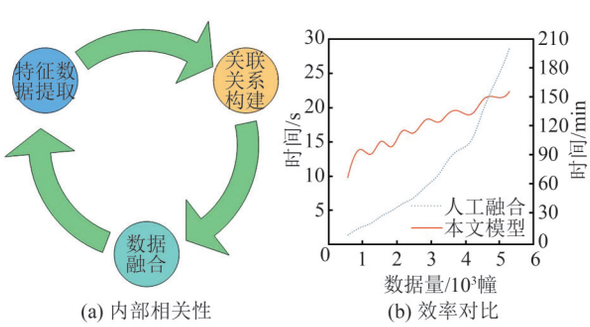


图3 主流算法效率对比

Fig. 3 Efficiency Comparison of Mainstream Algorithms



(a) 内部相关性 (b) 效率对比

图4 内部相关性与效率对比

Fig. 4 Internal Correlation and Efficiency Comparison

由图5可知,人工融合方法的重心在关联关系构建阶段,而本文模型的重心则在特征数据提取阶段,可见快速融合模型改变了多源异构数据融合流程的工作重心。尽管本文模型中的特征数据提取阶段会因此增加一定时间成本,但其成本增量可控,处于可接受范围内,而关联关系构建阶段所耗时间则会极大降低,进而降低整体流程所耗时间,并提高多源异构数据融合效率。

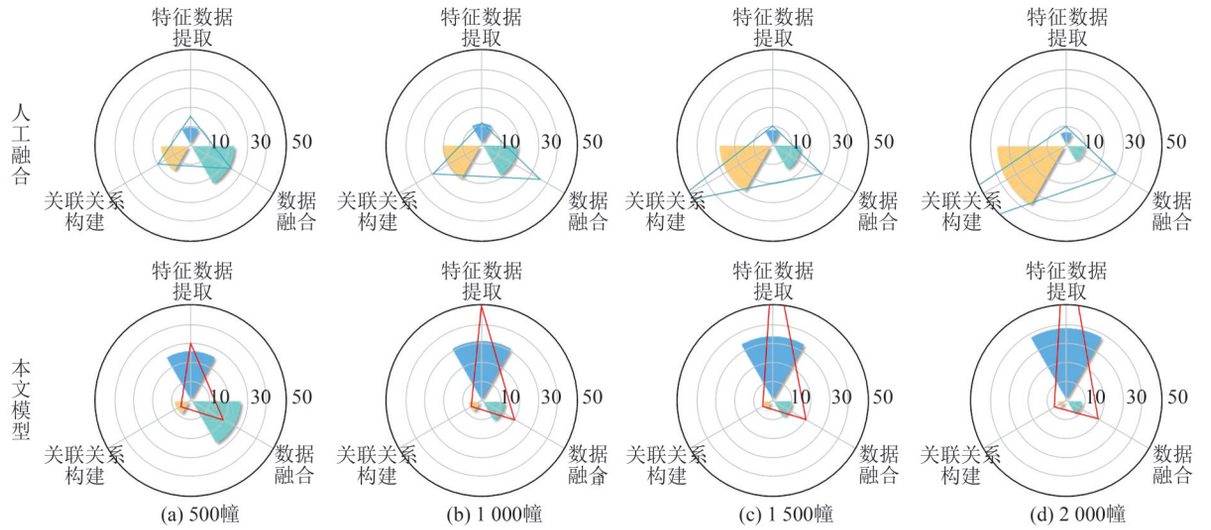


图 5　人工融合和本文快速融合模型效率对比

Fig. 5　Efficiency Comparison Between Artificial Fusion and Fast Fusion Model in This Paper

3　结　语

本文基于特征相似度提出了一种多源异构数据快速融合模型,该模型以相似度因子为核心,通过特征相似度匹配实现多源异构数据的批量融合。本文以不动产数据为例,对模型进行精度和效率的验证。经检验可知,模型融合准确率达到 93.50%,相较其他方法,拥有更为出色的性能,数据融合的可靠性更高。同时,相较其他方法,本文模型融合效率更高,可实现不动产数据的批量融合,每千幢的平均融合时间约为 1 s,提高了多源异构数据的融合效率,有助于提高不动产统一登记的效率,为不动产数据融合提供技术支撑。

参　考　文　献

[1] 李仕峰. 多源异构数据时空融合关键技术研究与应用[J]. 地理空间信息, 2021, 19(10): 19-21.
LI Shifeng. Research and Application of Key Technologies of Multi-source Heterogeneous Data Spatio-temporal Fusion[J]. *Geospatial Information*, 2021, 19(10): 19-21.

[2] DING Y, YU X, ZHANG J, et al. Application of Linear Predictive Coding and Data Fusion Process for Target Tracking by Doppler Through-Wall Radar[J]. *IEEE Transactions on Microwave theory and Techniques*, 2018, 67(3): 1244-1254.

[3] SUSPERREGI L, ARRUTI A, JAUREGI E, et al. Fusing Multiple Image Transformations and a Thermal Sensor with Kinect to Improve Person Detection Ability[J]. *Engineering Applications of Artificial Intelligence*, 2013, 26(8): 1980-1991.

[4] SHEN H, WU J, CHENG Q, et al. A Spatiotemporal Fusion Based Cloud Removal Method for Remote Sensing Images with Land Cover Changes[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(3): 862-874.

[5] LI K F. Smart Home Technology for Telemedicine and Emergency Management[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2013, 4: 535-546.

[6] ATAT R, LIU L, WU J, et al. Big Data Meet Cyber-Physical Systems: A Panoramic Survey[J]. *IEEE Access*, 2018, 6: 73603-73636.

[7] 涂伟, 曹劲舟, 高琦丽, 等. 融合多源时空大数据感知城市动态[J]. 武汉大学学报(信息科学版), 2020, 45(12): 1875-1883.
TU Wei, CAO Jinzhou, GAO Qili, et al. Sensing Urban Dynamics by Fusing Multi-source Spatiotemporal Big Data[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1875-1883.

[8] YU J, GU S, WANG J, et al. The Intelligent Decision-Making Based on Multi-source Heterogeneous Data Fusion in Manufacturing[C]//The 12th International Conference on Advanced Computational Intelligence, Dali, China, 2020.

[9] 莫慧凌, 郑海峰, 高敏, 等. 基于联邦学习的多源异构数据融合算法[J]. 计算机研究与发展, 2022, 59(2): 478-487.
MO Huiling, ZHANG Haifeng, GAO Min, et al. Multi-source Heterogeneous Data Fusion Based on Federated Learning[J]. *Journal of Computer Research and Development*, 2022, 59(2): 478-487.

[10] 夏伟,蔡文婷,刘阳,等.基于联合卡尔曼滤波的配电网多源异构数据融合[J].电力系统保护与控制,2022,50(10):180-187.
XIA Wei, CAI Wenting, LIU Yang, et al. Multi-source Heterogeneous Data Fusion of a Distribution Network Based on a Joint Kalman Filter[J]. *Power System Protection and Control*, 2022, 50(10): 180-187.

[11] 汤宇磊,吴杨杨,蒋兴征,等.面向自然资源信息提取的多源异构数据融合技术:以汉江流域NDVI数据为例[J].中国地质调查,2021,8(2):74-82.
TANG Yulei, WU Yangyang, JIANG Xingzheng, et al. Multi-source Heterogeneous Data Fusion Technology for Natural Resource Information Extraction: A Case Study of NDVI Data in Hanjiang Basin[J]. *Geological Survey of China*, 2021, 8(2): 74-82.

[12] LI X, WANG J, TAN J, et al. A Graph Neural Network-Based Stock Forecasting Method Utilizing Multi-source Heterogeneous Data Fusion[J]. *Multimedia Tools and Applications*, 2022, 81(30): 43753-43775.

[13] 魏之皓,贾克斌,贾晓未.多尺度特征融合的多源异构遥感数据水体提取[J].遥感信息,2021,36(5):41-48.
WEI Zhihao, JIA Kebin, JIA Xiaowei. Multi-scale Feature Fusion Based Multi-source Heterogeneous Remote Sensing Data Water Body Extraction[J]. *Remote Sensing Information*, 2021, 36(5): 41-48.

[14] YANG X, GUO W, ZHAO Z. A Multi-source Heterogeneous Data Fusion Method Based on Deep Learning[C]//The 4th International Industrial Informatics and Computer Engineering Conference, Sanya, China, 2019.

[15] 王宏伟,黄春林,侯金亮,等.基于数据同化算法融合多源数据估算雪深[J].武汉大学学报(信息科学版),2016,41(6):848-852.
WANG Hongwei, HUANG Chunlin, HOU Jinliang, et al. Estimation of Snow Depth from Multi-source Data Fusion Based on Data Assimilation Algorithm[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(6): 848-852.

[16] ZHANG Y, WANG Y, DING H, et al. Deep Well Construction of Big Data Platform Based on Multi-source Heterogeneous Data Fusion[J]. *International Journal of Internet Manufacturing and Services*, 2019, 6(4): 371-388.

[17] 戴海发,卞鸿巍,王荣颖,等.一种改进的多传感器数据自适应融合方法[J].武汉大学学报(信息科学版),2020,45(10):1602-1609.
DAI Haifa, BIAN Hongwei, WANG Rongying, et al. An Improved Multi-sensor Data Adaptive Fusion Method[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(10): 1602-1609.

[18] 张博,张猛,王非,等.VGI数据与地形图数据的自动融合研究[J].武汉大学学报(信息科学版),2019,44(11):1708-1714.
ZHANG Bo, ZHANG Meng, WANG Fei, et al. Automatic Data Integration Between VGI and Topographic Data[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(11): 1708-1714.

[19] LIU J, ZHANG X X. Data Integration in Fuzzy XML Documents[J]. *Information Sciences*, 2014, 280: 82-97.

[20] KHALILI A. An XML-Based Approach for Geo-semantic Data Exchange from BIM to VR Applications[J]. *Automation in Construction*, 2021, 121: 103425.

[21] 陈朝霞,熊骥华,张珊珊,等.基于FME的不动产单元测绘数据技术审查[J].测绘地理信息,2021,46(6):154-157.
CHEN Chaoxia, XIONG Jihua, ZHANG Shanshan, et al. Technical Review of Real Estate Unit Cadastral Data Based on FME[J]. *Journal of Geomatics*, 2021, 46(6): 154-157.

[22] 何思宇,赵孔阳.基于FME的土地利用现状图缩编技术研究[J].测绘与空间地理信息,2021,44(6):181-185.
HE Siyu, ZHAO Kongyang. Research on Downsizing Technology of Land Use Status Map Based on FME[J]. *Geomatics & Spatial Information Technology*, 2021, 44(6): 181-185.

[23] 李晨超.不动产登记数据整合要点探析[J].住宅与房地产,2017(21):20-21.
LI Chenchao. Analysis of the Real Estate Registration Data Integration Key Points[J]. *Housing and Real Estate*, 2017(21): 20-21.

[24] MCELHANEY R M. Algorithms for Graphics and Image Processing[J]. *Proceedings of the IEEE*, 1983, 71(9): 1116-1117.