引文格式：林尚纬,杜晓,张宏伟,等. 基于多源数据的全球路网融合生产应用研究[J]. 地理空间信息, 2025, 23(1): 90-94. doi:10. 3969/j.issn. 1672-4623.2025.01.019

基于多源数据的全球路网融合生产应用研究

林尚纬[1]，杜　晓[1]，张宏伟[1]，吴晨琛[1]，王艳东[2]，周　琦[1]，朱　玲[3*]，陈家阁[1]，万咏涛[1]
（1. 国家基础地理信息中心，北京 100830；2. 武汉大学 测绘遥感信息工程国家重点实验室，湖北 武汉 430079；
3. 中国农业大学，北京 100083）

摘　要：基于OpenStreetMap（OSM）等开源道路数据，设计了一套多源矢量道路数据获取与预处理、要素匹配融合、名称翻译和属性挂接的生产加工技术流程，从多源路网几何信息和属性信息两个方面进行获取挖掘和补充。对乌拉圭进行道路融合生产应用的结果表明，相较于初始OSM道路，匹配融合加工后的道路几何增长率约为20.31%，准确率约为97%，属性整体增长率约为4.64%，新增属性约为8.91%，准确率约为93%。相关研究成果已应用于境外地理信息资源建设等工程中，可为全球范围核心矢量要素数据产品制作和更新提供借鉴依据。

关键词：自发地理信息；OSM；几何匹配；数据融合；矢量要素；路网
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2025）01-0090-05

Research on the Application of Global Road Network Fusion Production Based on Multi-source Data

LIN Shangwei[1], DU Xiao[1], ZHANG Hongwei[1], WU Chenchen[1], WANG Yandong[2], ZHOU Qi[1], ZHU Ling[3], CHEN Jiage[1], WAN Yongtao[1]
(1. National Geomatics Center of China, Beijing 100830, China;
2. State Key Laboratory of Information Engineering in Surveying, Mapping & Remote Sensing, Wuhan University, Wuhan 430079, China;
3. China Agricultural University, Beijing 100083, China)

Abstract: Based on open-source road data such as OpenStreetMap (OSM), we designed a production technology process for multi-source vector road data acquisition and preprocessing, feature matching fusion, name translation and attribute linking. We obtained, mined and supplemented the geometric information and attribute information of multi-source road network. The application of road fusion production in Uruguay shows that compared to the initial OSM road, the geometric growth rate of the road after matching fusion processing is about 20.31%, the accuracy rate is about 97%, the overall growth rate of attributes is about 4.64%, the new attribute is about 8.91%, and the accuracy rate is about 93%. The relevant research results have been applied in overseas geographical information resource construction and other projects, providing reference for the production and update of core vector element data products on a global scale.

Key words: volunteered geographical information, OSM, geometric matching, data fusion, vector element, road network

《自然资源部信息化建设总体方案》提出，提升基础地理信息服务保障能力，加强全球地理信息资源建设，有效解决我国自主全球多尺度地理数据产品匮乏、覆盖全球高精度自主位置服务能力不足的问题[1]。因此，境外地理信息资源数据建设成为一个迫切需求，尤其是对于道路这种精度要求高、更新频率快的矢量数据进行境外地理信息资源数据采集存在相当大的难度。随着空间数据获取与处理技术的迅猛发展，矢量空间数据日益增多[2]，自发地理信息（VGI）开源数据可作为全球范围道路数据生产的首选数据源。以OpenStreetMap（OSM）为代表的开源数据具有获取免费、现势性较强等优点，诸多学者对其数据质量进行了验证，如朱富晓等[3]通过构建多层次多粒度模糊综合评价模型实现了宏观与微观相结合、全局与局部相结合的OSM数据质量多维度多层次综合度量；王艳东等[4]得到了OSM路网距离、角度几何精度的评价结果；罗路长[5]、刘波等[6]从长度完整性和道路名称完整性两个方面对其数据质量进行了对比分析；刘纪

项目来源：国家重点研发计划资助项目（2022YFC3802903）。
第一作者简介：林尚纬（1992—），硕士，工程师，研究方向为泛在地理信息数据获取、融合和GIS空间分析，E-mail：linshangwei@ngcc.cn。
通信作者：朱玲（1994—），硕士，研究方向为大数据与人工智能技术在地理信息、农业、渔业及自然资源等方面的应用，E-mail：zhulingcau@cau.edu.cn。

平等[7]分析互联网泛在地理信息发现，公众VGI具有位置信息完整度较高、信息主体数量较多、自动更新等特点；Haklay M[8]对比英国OSM数据与权威地形测量数据发现，其高速公路重合率达到了80%。以OSM为代表的众源数据因其获取成本低、现势性强等优点成为全球地理信息资源建设的重要数据源[9]。

由于OSM数据的采集者主要是广大公众志愿者，因此在人口流动较大的部分区域，数据长度的完整性较高，比同一时间精度下的百度地图等互联网地图更详细、道路类型更丰富，道路的现势性也较高[10]。OSM数据作为主要的境外全球矢量数据来源，具有丰富的几何信息，但区域性差异较大，且属性信息的完整度较低[11]。近年来利用深度学习从遥感影像上自动或半自动提取路网数据的技术得到了较快发展，在一定程度上降低了提取路网的难度，解放了生产力[12]，弥补了开源数据空间形态缺失的问题，可作为OSM几何信息的补充渠道；同时开源地理空间数据提供了丰富的语义信息，如互联网地图提供包含全球城市政区、交通和商业信息的矢量地图，具有丰富的POI信息，GeoNames集成了不同语言的地名、人口数量、更新时间等地理信息数据，也可作为OSM属性信息的补充渠道；此外，利用VGI数据进行道路匹配和融合的方法研究也取得了一定的进展[13]。本文设计了一套多源路网矢量数据获取转换、几何匹配融合、地名翻译和属性挂接的全流程生产加工技术方法，以OSM路网数据为基础，以遥感影像智能提取几何路网和大数据挖掘地名地址信息为补充，结合矢量要素预处理、几何匹配、融合、地名翻译和属性挂接方法，生成一套几何和属性信息更加丰富且具有中外文道路名称的全球路网标准数据产品。

1　研究思路和方法

1.1　研究思路

基于多源数据的全球路网融合生产技术流程为：在确定生产任务区范围后，首先获取开源OSM数据，通过自定义分类转换模板（道路、水系等）分别提取道路几何要素及其对应属性表；再通过深度学习等方法提取高分辨率遥感影像上的增量道路几何要素，利用互联网大数据挖掘含有道路名称、坐标等信息的属性数据，并经过格式转换、清洗、分类、去重等预处理操作，得到任务区范围内的增量道路几何和属性数据；然后将OSM道路几何要素与增量道路几何要素进行要素关联、匹配、融合，并与已有的高精度数字正射影像（DOM）进行几何配准、拓扑编辑加工整理，生成几何要素融合成果，同时将OSM道路名称属性表与增量道路名称属性表进行匹配、去重、翻译，获得道路名称属性融合成果；最后通过唯一字段标识或空间关系匹配赋值等完成道路属性至道路几何要素的属性挂接，并通过DOM位置协同、精度验证等质量检查生成任务区最终的路网数据产品。

1.2　众源路网几何要素获取

1.2.1　OSM道路几何要素获取

OSM地图的矢量数据以开放数据库授权方式授权，数据包括道路与铁路路网、建筑、水体、土地利用、兴趣点、行政区边界等。本文通过自主开发的OSM数据转换工具实现对OSM中路网数据的分类分级转换提取。

1）数据下载。OSM提供了按大洲、国家、自定义范围3种下载方式，确定任务区范围后，可通过OSM官网或GeoFabrik下载服务等链接下载原始数据，为便于获得更多要素属性字段信息，本文选取含有全部属性字段的osm或pbf格式。

2）分类模板制定。引入图层分类模板对OSM数据进行分类，图层分类模板采用XML文件表示，记录图层与OSM中TAG属性间的对应规则。第一级标签Category，记录模板名称；第二级标签Layer，记录图层名称、别名和要素类型，每个Layer为一个类别；Layer下设置Fields标签，用于存放每个图层包含的属性字段。

3）数据转换。将下载的OSM原始数据按照制定的分类模板，提取相应类别的Shapefile或File Database类型矢量几何数据以及包含每条要素所有字段信息的Excel属性表。

1.2.2　增量路网自动提取

利用深度学习遥感影像信息提取技术，通过标签更新、样本制作、模型制作评价等步骤，利用深度学习模型和数据自动提取道路几何要素。基于MapBox、OSM等开源矢量瓦片解译规则，通过爬取互联网地图服务矢量瓦片、解析其数据结构、设计矢量瓦片转换方法、构建几何要素重构模型重建要素几何信息，得到增量路网数据。

1.3　多源道路属性信息获取

针对OSM路网和增量路网数据中部分道路属性缺失的问题，本文以下载转换互联网公开地名数据库资源补充为主，以大数据爬取百度、天地图等互联网地图服务、多源POI道路地名地址解析为辅的方式挖掘补充道路属性信息。

1.3.1　公开地名数据库

对于结构化和半结构化数据，通过面向地名实体的通用搜索、聚焦搜索和深度搜索技术，实现地名数据的发现与获取；再通过要素分类模板实现地名数据

匹配与属性抽取。在OSM几何路网自带属性信息的基础上，选取GeoNames作为道路属性的主要补充数据源。GeoNames由Marc Wick及其团队创建，是全球范围开放共享的数据资源。GeoNames地名数据被划分为9个特征类，其中R代表交通地名。首先下载任务区原始txt格式的GeoNames数据压缩包，并将txt导入Excel表格；再根据分类体系说明筛选保留交通类地名，记录名称、经纬度坐标等属性内容。在ArcGIS等地理信息软件中导入上述带有经纬度坐标的GeoNames交通地名数据，显示 X、Y 数据并导出为交通地名点要素矢量数据。

1.3.2 泛在地名挖掘

对于非结构化数据，采用网络搜索、矢量瓦片要素重建、OCR文字识别等技术获取含有地名属性和地理坐标的网络数据，再对其进行地址解析、格式转换、筛选、属性抽取、分类、异常值删除等处理，形成统一格式的增量道路名称属性数据集。各类地图服务为用户提供了丰富的POI数据，根据地址信息间的语义相似度进行聚类，划分含有不同道路名称的地址信息，再采用基于词聚类的道路名称抽取方法，从多源POI数据的属性信息中提取道路名称。

1.3.3 地名数据预处理

对多源道路属性数据进行编码、坐标、格式和数据结构等转换，再经过特殊字母去符号化、统一字母大小写、消除无意义干扰字符等步骤，采用顾及别名的匹配算法[14]进行多源道路属性融合匹配，剔除无效、重叠、错误的属性内容，最大限度地保留有效属性信息，确保数据的一致性和可靠性。

1.4 要素几何匹配融合

1）预处理。将OSM初始路网和增量路网数据集的数据格式转换为ArcGIS等软件通用的shp线要素格式，空间参考统一为GCS_WGS_1984坐标系，对每条线要素进行整合、融合、打断相交线等操作，以消除悬挂点、伪节点等拓扑错误，提高后续路网几何要素的匹配效率和准确度。分别遍历数据集中每条线要素在其缓冲区范围内与同一数据集中的其他要素之间的坐标位置关系：当两条线要素的折点间距离在指定容差范围内，且要素属性相同或仅一条有属性时，将折点赋予相同的坐标位置（两折点连线中点），并重新连接两条几何要素；当一条线要素的折点位于任何其他线要素的容差范围内时，则在其他线要素边上插入新折点并延伸连接至新折点，整合后效果见图1。当两条线要素具有一个公共折点或在容差范围内，且具有相同的道路等级和属性信息或仅有一条有属性时，将两条线要素融合为单个线要素，融合后效果见图2。在道路相交处进行添加新增公共折点打断处理，将线要素分割成多条线要素，同时移除叠置线段。

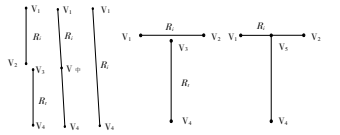


图1 整合示例图

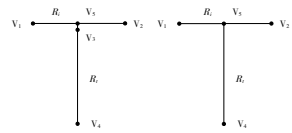


图2 融合示例图

2）几何匹配融合。本文采用一种综合中值Hausdorff距离MM _HD[15]、方向角差值Angle、重叠度Over的道路几何多因子匹配方法进行要素实体匹配，建立要素与要素之间的关系，并根据道路要素的优先级逐级融合。首先遍历OSM初始数据集中每条线要素与其一定范围内的增量路网数据；再对增量路网中的待匹配对象进行加权计算获得其置信度，若置信度大于预设阈值，则将该线要素与OSM初始数据集进行融合；最后输出几何道路融合数据集，并对其再次进行要素整合、融合、打断相交线等操作，以消除融合后道路结果中的拓扑错误，提高资料使用效率。Angle因子通过计算两条线要素的最小面积外接矩形对角线方位角之差，并取其绝对值进行判断。Over因子通过计算以两条线要素之间MM_HD距离为半径建立的缓冲区交集面积与并集面积的比值进行判断。

$$\mathrm{CON}_i = w_1 \times \mathrm{MM_HD}' + w_2 \times \mathrm{Angle}' + w_3 \times \mathrm{Over}'$$

式中，CON_i 为增量路网中第 i 条待匹配道路的置信度；w_1、w_2、w_3 分别为MM_HD、Angle和Over的权重，通过层次分析法求得；$\mathrm{MM_HD}'$、Angle'、Over' 分别为经过Min-Max归一化后的值。

1.5 地名翻译和属性挂接

全球范围道路数据属性存在多语种情况，在项目工程实际生产中分为道路数据制作和道路名称翻译（外文翻译为中文）两个环节，通常分别由测绘生产单位和外文翻译单位独立完成，因此需要预先将道路中含有外文名称的信息提取出来，并记录其要素唯一编码或经纬度坐标，待名称翻译和数据生产完成后，在建库阶段对二者进行属性挂接。道路名称翻译采用基于众源地名数据的自动化翻译方法[14]，通过构建地名翻译规则库和语料库，遵循专名音译、通名意译的原则，对转换提取的OSM地名和增量挖掘地名进行快速融合匹配、规模化翻译生产。属性挂接分为两种方式：①将翻译后的OSM道路属性信息与几何要素融合成果中具有OSM唯一ID标识字段的几何道路进行关联匹配和属性挂接；②将翻译后的多源补充道路属性信息与几何要素融合成果中所有道路通过最邻近策略（道路要素外扩一定距离范围）进行关联匹配和属性赋值，当道路要素具有OSM属性或自带属性信息时，

保留原始属性信息，不进行替换赋值。

2　生产实验与结果分析

2.1　实验区概况

本文选取南美洲乌拉圭作为实验区。乌拉圭地处南美洲东南部，北邻巴西、西接阿根廷、东南濒大西洋，国土总面积约为17.62万km^2，是南美洲最小的国家。乌拉圭交通发达，道路平坦宽阔，公共交通较便利，陆路交通以公路运输为主。

本文采用的行政区划边界数据为GADM的乌拉圭初始几何属性数据（osm格式），时间为2022年1月；增量道路几何数据来源于互联网地图矢量瓦片挖掘和基于ZY-3等高分辨率影像的自动提取；多源道路属性信息来源于GeoNames地名数据库和互联网地图OCR文字识别等。

2.2　数据预处理与匹配融合

根据OSM交通要素层的分类内容，设置道路层分类模板和查询条件标签<Exist Key="highway">，在下载的任务区OSM数据中查找相应标签中的键值(Key, Value)，通过数据转换工具提取转换高速公路、干线公路、主要公路、次要公路、三级公路等几何矢量要素，同时从数据标签中提取带有名称Name的道路名，记录当地语言名、英文名及其所在要素唯一ID号，整理形成OSM道路名称记录表。对获取的OSM几何要素和增量几何矢量要素参照表1的对应关系进行道路等级映射，形成全球统一的道路分类等级。

表1　道路等级分类映射对照关系

几何要素类别	中文类别	全球道路分类代码（CLSID）	几何要素类别	中文类别	全球道路分类代码（CLSID）
motorway	高速公路	12001	escape	紧急避险道	12012
trunk	干线公路	12002	ford	徒涉场	12013
primary	主要公路	12003	raceway	赛车道	12014
secondary	次要公路	12004	bus_guideway	引导性公交线	12015
tertiary	三级公路	12005	cycleway	自行车道	12016
motorway_link	高速公路匝道	12006	pedestrian	步行街	12017
trunk_link	干线道路匝道	12007	footway	步道	12018
primary_link	主要道路匝道	12008	steps	台阶路	12019
secondary_link	次要道路匝道	12009	track	小路	12020
tertiary_link	三级道路匝道	12010	trail	小路	12020
unclassified	未分级的道路	12011	bridleway	马道	12021
residential	通往住宅区的道路	12011	path	特殊道路	12022
service	辅助道路	12011	road	未知道路	12023
living_street	居住区街道	12011	...	...	...

对获取的多源道路属性数据集进行地名预处理，遍历其与OSM几何要素，通过判断位置与属性关系进行第一次去重处理，判断范围设置为10 m，处理原则为仅保留OSM道路中Name为空且在10 m阈值范围内的多源道路属性数据记录；当二者都有Name属性值时，以OSM道路的Name为准，不保留多源道路属性数据集中对应的属性记录。为便于后续的数据匹配和属性挂接，属性数据集均采用UTF-8编码数据格式。UTF-8是一种包含全世界所有国家所用字符的国际编码，可消除不同软件特殊字符的乱码问题[16]。

对获取的众源路网几何要素数据集进行匹配融合，生成几何要素融合成果，并将处理后的多源道路属性数据集和OSM道路名称记录表一并下发给专业地名翻译单位进行翻译，通过位置最近邻匹配策略和OSM唯一ID标识字段关联挂接两种方式，对翻译后含有中外文名称属性的道路属性数据集与几何要素融合成果的名称属性进行挂接赋值。

2.3　实验结果分析

2.3.1　道路几何分析

对比分析几何匹配融合前后以及经生产单位作业人员加工处理和质量验收后的成果数据，乌拉圭初始OSM几何道路要素为116 866条，总长度为97 630 km，补充增量道路要素为95 200条，长度为19 832 km，道路长度提升率约为20.31%。要素几何匹配融合前后部分地区道路叠加谷歌高分辨率卫星影像的效果见图3、4。



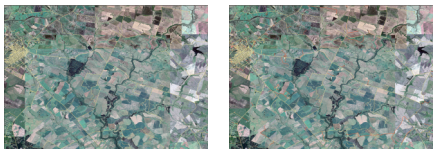
a 初始几何道路　　b 融合后道路

图3　沿海地区融合对比

注：黄色、红色分别表示OSM初始几何道路和补充道路。

融合后的道路数据集经过中外文翻译、生产单位

加工处理、属性挂接和成果质量验收后，最终汇交道路成果数据要素275 761条，总长度为98 631 km。进一步分析发现，经过整合、融合和打断操作后，道路要素总数明显多于初始OSM和补充增量道路数据集之和；汇交道路成果数据中道路总长度比OSM初始道路增加了约1 000 km，仅占补充增量道路要素总长度的5%，补充增量道路数据利用率较低。全球项目核心矢量要素数据生产技术规程对OSM以外的新增道路补充采集的技术指标要求为一般选取路宽大于20 m（含20 m）且通达较远（长度大于1 km），同时连接三级及以上道路或乡镇级及居民地间的主要车辆通行的道路。由于乌拉圭境内OSM初始道路数据现势性较强、主要道路覆盖完整，补充增量道路要素多为等级较低、宽度较窄、长度较短的道路，长度和宽度等难以满足技术指标要求，且部分增量道路在DOM产品中不支持，因此在实际工程生产中按照采集要求舍弃了较多不满足技术指标的增量道路。本文按照1‰的比例随机抽样对补充增量道路要素进行人机交互检查，检查几何位移、拓扑接边等位置精度和逻辑一致性，道路几何补充准确率约为97%。



a 初始几何道路　　b 融合后道路

图4 内陆地区融合对比

注：黄色、红色分别表示OSM初始几何道路和补充道路。

2.3.2 道路属性分析

统计分析发现，乌拉圭初始OSM几何道路要素中有名称属性的要素为54 640条，总长度为36 650 km，占比约为37.54%，补充道路属性信息为12 876条，补充道路属性既可补充原始OSM中无道路名称的要素，又可补充增量道路的几何要素属性；经过中外文翻译、生产单位加工处理、属性挂接和成果质量验收后，最终汇交道路成果数据要素中有名称属性的要素为180 106条，长度为41 605 km，占比约为42.18%，道路属性整体提升率约为4.64%。进一步分析发现，去重后OSM道路名称记录表为15 283条，补充道路属性信息与OSM道路属性信息表去重后剩余1 362条，可得到新增道路属性信息提升率约为8.91%。本文按照1%的比例随机抽样对补充道路属性信息进行人机交互检查，检查分类代码值、属性赋值等属性精度和逻辑一致性，道路名称属性补充准确率约为93%。最终汇交道路成果数据通过了任务承担单位的两检一验以及国家测绘产品质量检验测试中心的成果质量核查，成果质量优良品率大于85%。最终道路数据配图效果和属性表见图5和表2。



图5 道路配图效果

表2 道路属性表

FID	Name	CNAME1	CLSID
0	Ruta 18	18国道	12002
22	Rincón	林孔路	12011
43	Ruta 26 Brigadier General Leandro Gómez	26国道莱安德罗戈麦斯路	12003
57	Caballero	卡瓦列罗路	12011
96	Yuri Gagarin	尤里加加林路	12011
121	Puente Internacional Barón de Mauá	毛阿男爵国际桥	12005
128	Apolinario Rivero	阿波利纳里奥里韦罗路	12011
132	Pedro Figari	佩德罗·菲加里路	12011
162	A Blanca Diaz	A·布兰卡·迪亚兹路	12020
176	Joaquin de Paz	华金德帕斯路	12011
185	Guaraní	瓜拉尼路	12011
224	Joaquín Ramón Travieso	华金拉蒙特拉维索路	12011
234	Ámsterdam	阿姆斯特尔达姆路	12011
253	Ituzaingó	伊图萨因戈路	12011
260	Joaquín de Paz	华金德帕斯路	12011
316	Ruta 9 Coronel Leonardo Olivera	9国道莱昂纳多·奥利韦拉	12002

3 结 语

本文结合测绘工程项目实际生产流程，构建了基于多源数据的全球路网融合生产技术体系，从多源路网几何信息和属性信息两个方面进行挖掘和补充；并对境外多语种属性信息的翻译和匹配挂接流程进行详细梳理，以OSM为基本数据源，选取境外部分区域进行了工程试生产验证。结果表明，本文的技术方法能在一定程度上补充境外众源路网数据的几何和属性信息，可为全球范围核心矢量要素数据产品制作和更新提供借鉴依据。未来将探索其他多源矢量要素（如水系、建筑物、耕地等）的融合生产应用，进一步完善本文的研究思路方法。

参考文献

[1] 吴洪涛,张子平,贾萍,等. 自然资源信息化面临的挑战与实施策略:从《自然资源部信息化建设总体方案》说起[J]. 北京规划建设,2020(增刊1):10-15

（下转第122页）