

doi:10.11835/j.issn.1000-582X.2017.09.010

自发地理信息的应用问题及应对策略

杨清丽¹, 马超², 徐青²

(1.61243 部队, 兰州 730020; 2.信息工程大学 测绘学院, 郑州 450052)

摘要:自发地理信息的迅速发展逐渐成为了地学领域的研究热点,研究分析了自发地理信息应用于地图制图更新中面临的数据质量评价、几何数据融合与属性信息匹配转换等核心问题,在此基础上提出了自发地理信息分析与融合处理的一般流程,并针对每个问题分别提出了相应的解决策略;最后,利用北京地区 OSM 数据,对流程和策略进行了验证。

关键词:自发地理信息,数据质量评价,数据融合,属性匹配

中图分类号:TP393.02

文献标志码:A

文章编号:1000-582X(2017)09-083-06

The application problems and coping strategies of volunteered geographic information

YANG Qingli, MA Chao, XU Qing

(1.Unit 61243, Lanzhou 730020, P.R.China; 2. School of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, P.R.China)

Abstract: The rapid development of volunteered geographic information generally becomes research focus in the field of geoscience. This paper analyzes the core questions of the evaluation of data quality, the fusion of geometric data and the conversion and matching of attribute information. during the volunteered geographic information used in mapping update, then proposes a general process of volunteered geographic information analysis and fusion operation, and finally puts forward the corresponding solutions for each question. Finally, the process and the strategy was verified by using Beijing OSM data.

Keywords: volunteered geographic information; data quality evaluation; data fusion; attribute matching

地理信息技术的发展始终和计算机技术、网络技术息息相关。以个人为中心的互联网时代,逐步形成了用户创建内容(UGC, user generated content)的概念,为各种互联网应用带来全新的技术和理论支持,也为地理信息网络服务带来了新的机遇,自发地理信息(VGI, volunteered geographic information)的概念应运而生^[1-4]。VGI是指用户通过在线协作的方式,以普通手持GPS终端、开放获取的高分辨率遥感影像以及个人空间认知的地理知识为基础参考,创建、编辑、管理、维护的地理信息^[5]。与传统地理信息采集和更新方式相比,来自非专业大众的VGI数据具有现势性高、传播快、信息丰富、成本低、质量各异、分布不均匀等特点,逐渐成为近年来国际地理信息科学领域的研究热点,可应用于应急制图、交通分析、早期预警、地图更新、犯罪

收稿日期:2017-01-20

基金项目:国家自然科学基金资助项目(41571399)。

Supported by National Natural Science Foundation of China(41571399).

作者简介:杨清丽(1976—),女,重庆大学硕士,主要从事地图制图及GIS应用工作,(Tel) 15339830468;
(E-mail)3472771325@qq.com。

分析、疾病传播分析等诸多地理空间信息服务领域^[6-10]。

但是,VGI 数据由于数据质量不明确、数据分布不均匀、属性信息不统一等问题,还不能直接应用于地图制图与更新中。为此,在深入分析 VGI 数据与现有矢量数据融合过程中面临的数据质量评价、几何数据融合与属性信息匹配转换等核心问题的基础上,针对这些问题提出了相应的解决策略,为 VGI 的地图制图与更新应用构建了分析处理流程。

1 VGI 用于数据源存在的问题

由于目前地理信息数据生产以矢量数据为基础,因此,研究对象主要是指 VGI 矢量数据。与专业地理信息相比,VGI 数据无论是数据来源、采集方式,还是数据质量、权威性等方面,均与现有专业地理信息有所不同。

1.1 数据质量未知

空间数据由于空间的复杂性、模糊性、人类认识和表达能力的局限以及人们在数据处理过程中的人为因素,总是与真实值之间存在一定的差异,这种差异不能完全消除,这就是数据质量问题^[11-13]。VGI 原始数据是指用户上传的数据,需要经过 VGI 平台的处理后提供给其他用户。VGI 数据质量问题为:1)用户没有专业测绘知识,经过手机或者车载 GPS(GPS 精度不够或者不同的设备精度差异)、在遥感影像上勾勒得到数据,具有人为主观性,在测绘数据关注于属性数据的规范性,字长和数据类型都有认知局限,也给地物的属性信息的采集带来了很大困难。2)VGI 数据可能存在恶意内容,数据冗余量大、容易泄露用户的隐私内容等。

1.2 几何数据细节繁杂

一般情况下,VGI 数据的来源包括志愿者上传的 GPS 轨迹、依据高分辨率遥感影像采集的数据、商业公司和政府部门捐赠的数据等,因此,VGI 数据具有无尺度的特点,数据细节丰富、详实。如图 1 所示

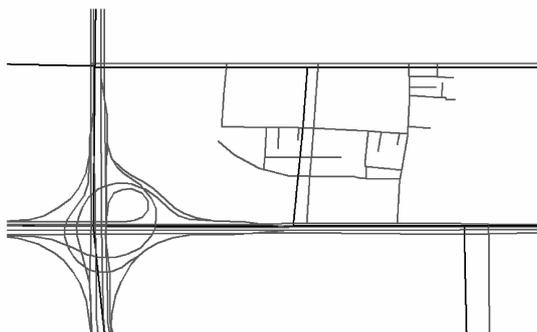


图 1 VGI 数据与导航数据对比

Fig 1 VGI data and navigation data comparison

由图 1 可以看出,由于 VGI 数据由志愿者上传的轨迹数据处理而成,等同于实测数据,故此造就了 VGI 几何数据细节十分丰富,例如在 OSM 数据中,主干道一般都呈现为多条平行路(2 条、4 条,甚至更多),主干道的交叉口的复杂立交结构、各种复杂的道路交叉口等微观结构也被表示出来。这些丰富的数据细节由此也成为了数据处理的巨大负担,制约了数据生产的速度和效率,需要进行大量的选取、化简等制图综合处理。

1.3 属性信息不一致

许多行业 and 部门对空间数据的需求日趋强烈且多样化,纷纷建立具有各自特点的空间数据库。由于各自在建库过程中采用不同的数据模型设计,对语义表达上不一致,主要表现要素的分类分级不同、采用不同的属性字段描述相同的地理实体、相同含义相同类型的属性字段取值不同。以道路的分类为例,国家基础地理信息数据与 VGI 数据的道路分类如表 1 所示。

表 1 不同数据道路分类分级比较
Table 1 Comparison of classification of different data road classification

OSM 道路分级	国标道路分级	日本地图册道路分级
主要道路	国道	国道
小路	省道	主要都府道
连接道路	县道	3 m 以上道路
微小道路	乡道	2 m 以上道路
未知类型道路	主要街道	1 m 以上道路
—	次要街道	小径

由于 VGI 数据是来自于全世界不同国家和地区的公众共同创建和维护,这些人的文化背景、风俗人情各不相同,由此导致 VGI 数据属性信息杂乱,存在各种不一致。此外,没有统一的标准,也是造成 VGI 语义信息杂乱的重要原因。以 OSM 数据为例,该数据属性信息以“key = value”的方式出现,用户除了可以选取已有的标签以外,还可以根据需求自定义标签,由此造成了属性信息的严重不统一。这种开放性的标签规则,导致了属性数据的极大冗余,根据官方统计数据,目前该数据已有超过 40 000 种数据标签,为语义分类编码的统一带来了严峻的挑战。此外,由于不同国家、地区、来源等数据的属性的分类、分级均不相同,也阻碍了 VGI 数据与现有数据的属性信息匹配与转换。

2 VGI 数据分析与融合处理技术框架

通过对 VGI 数据存在的问题剖析,针对矢量空间数据生产与更新应用需求,构建了相应的 VGI 数据分析与融合处理流程(见图 2)。

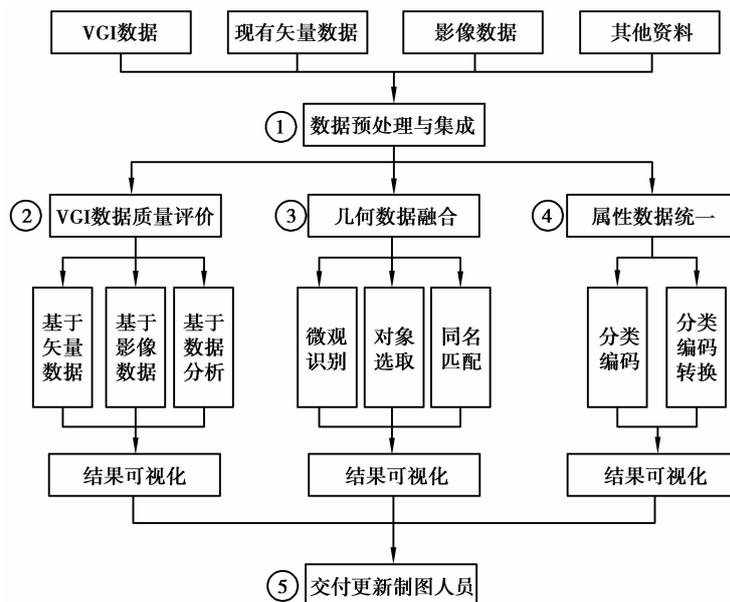


图 2 VGI 数据分析与融合处理流程

Fig 2 VGI data analysis and integration process

2.1 数据预处理

进行数据预处理主要是为了将 VGI 数据与现有矢量数据、参考资料等进行集成,包括数据格式的转换、

数学基础转换等。以 OSM 数据为例,其数据格式主要有 Shapfile 矢量数据格式和.osm 自定义的数据格式,但是目前数据生产格式为国家标准数据格式,需要将 OSM 数据转换成为国家标准数据格式,方能为其集成提供基础。其次是数学基础转换,OSM 数据采用了 WGS84 大地坐标系,没有进行投影,而国标数据格式采用的是中国 2 000 大地坐标系,因此,必须进行 2 种大地坐标系统之间的投影转换。

2.2 VGI 数据质量评价

为了对 VGI 数据质量进行科学、合理的评价,笔者提出了 3 种评价策略。1)基于矢量数据的评价方法。主要以已知高质量的矢量参考数据作为对比对象,将 VGI 数据与之进行叠加,并依据事先约定的质量因子,逐一进行对比、计算。这些质量因子包括几何定位精度、属性精度、几何完整性、属性完整性、主题一致性、拓扑关系一致性等。2)基于影像的评价方法,是以经过几何纠正的影像数据作为参考对比对象,将 VGI 数据与之进行叠加,并通过几何纠正的方法,分析计算 VGI 数据的几何定位精度。但是由于影像数据只有几何信息,没有属性信息,因此,无法进行属性信息的评价。3)基于数据可信度的评价方法。在没有参考数据的情况下,提出了基于数据可信度的评价策略,将 VGI 数据质量转换为数据可信度,通过可信度来反映 VGI 数据质量。如基于点数统计的志愿

者信誉度计算公式如式(1)所示

$$R_i = \sum P_s / \sum P_a, \quad (1)$$

式中: R_i 表示用户 i 的信誉度; $\sum P_a$ 表示用户上传的所有点; $\sum P_s$ 表示用户所有上传点中被保留的点。

2.3 几何数据融合

将 VGI 与已有矢量数据进行集成与融合处理是进行数据生产与更新的核心环节,最主要的是几何数据的融合。以 OSM 数据的道路网数据处理为例:首先识别、化简道路网数据中的微观结构,包括天桥数据、复杂交叉口等微观结构,这些道路微观结构大大增加了与现有矢量数据的集成与融合的难度;其次,需要对道路网进行选取处理,由于不同尺度地图数据的承载量不同,因此需要根据需要对道路网数据进行选取处理,选取重要性程度较高的道路;最后,将经过化简、选取处理的 OSM 道路网数据与现有道路数据进行同名实体匹配处理,以便于制图人员发现变化,进而对道路网进行更新操作。

2.4 属性数据统一

VGI 数据开放性的获取方式,导致了属性信息存在着各种不一致,需要与已有数据的分类分级与编码进行转换统一。以 OSM 数据为例,首先进行其数据本身的分类分级统一,由于 OSM 数据用户可以自定义属性数据标签,因此,相同等级的道路可能会存在多种不同的标签进行描述,在进行属性转换之前,首先要对这些标签进行语义匹配与统一;其次,将统一语义后的 OSM 数据分类分级与已有数据的分类分级进行统一,在分类分级标准不同的情况下,采用基于控制文件的属性特征映射与转换策略;而在分类分级标准相同的情况下,则采用基于同名实体匹配的属性特征合并与更新策略。

2.5 交付制图人员

将上述经过融合处理的 OSM 数据,提交给制图人员,进行后续的数据编辑、加工、删选等数据生产工作。

3 实例验证

以北京主城区 OSM 道路数据进行实验,主要验证以上流程和策略的可行性。数据来源于 OpenStreetMap 项目的历史数据库,包括主要道路、高速公路、街区小路、步行道路等几何数据(暂时不考虑属性数据)。首先进行 OSM 数据预处理,由 WGS84 大地坐标系,转换为中国 2 000 大地坐标系,即投影转换;其次,进行数据质量评价,为了便于分析比较,以北京四维图新 2014 年出版的导航数据作为参考数据,与实验数据进行匹配比较;再次,进行几何数据的化简融合,选取重要性程度较高的道路、摒弃一般意义的支

路、小路等;最后,将经过化简、选取处理的 OSM 道路网数据与现有道路数据进行同名实体匹配处理,提取出可用于更新的道路数据,制作交通图。表 2 为依据公示(1)计算的部分按瓦片划分实验区域数据质量,图 3 为部分区域的实验结果,深色道路为化简后的主要道路。本文提出的 VGI 数据分析与融合框架与策略,可以提高空间矢量数据的生产质量和生产效率,解决当前 VGI 数据应用中的部分困难。在后续的研究中,应进一步建立属性数据的对应转换关系,探索更广泛区域的 VGI 数据分析与融合处理方法。

表 2 实验区域数据质量统计

Table 2 The statistics of the data quality of experiment area

编号	完整性	精度	质量
1	0.83	0.56	0.695
2	0.79	0.62	0.705
3	0.92	0.73	0.825
4	0.69	0.66	0.675
5	0.87	0.59	0.730
6	0.84	0.82	0.830
7	0.93	0.89	0.910
8	0.74	0.71	0.725
...
35	0.81	0.64	0.725
36	0.68	0.71	0.695

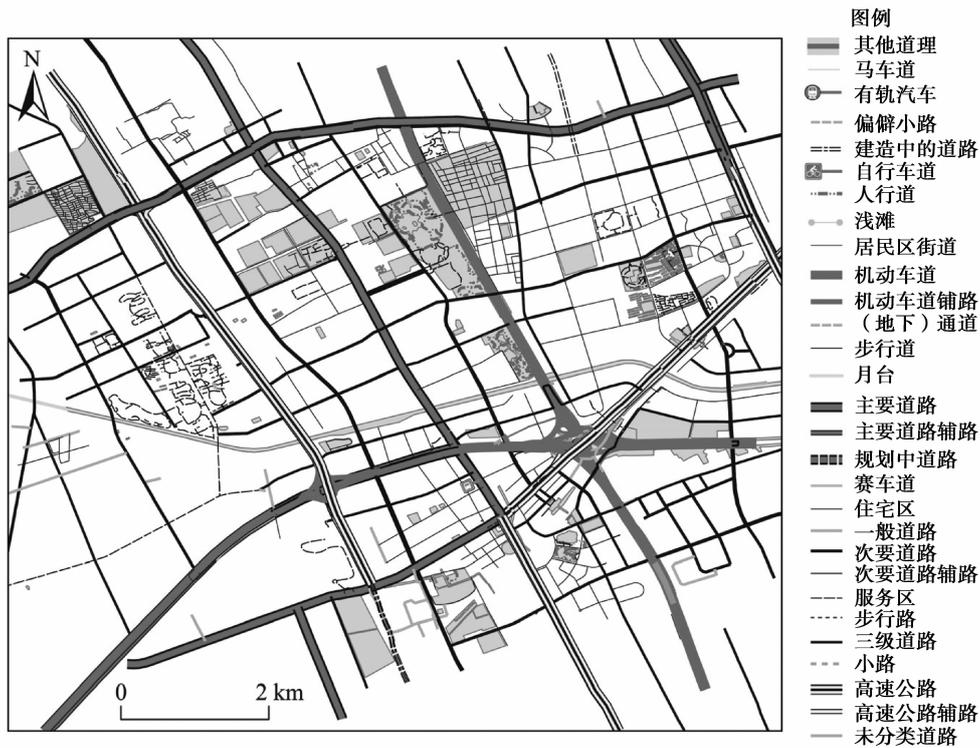


图 3 部分实验区域交通图

Fig.3 Part of the experimental area traffic map

4 结 语

VGI数据将是未来重要的地理信息获取方式之一,但是目前VGI数据由于受到数据质量等诸多因素的影响,还未能得到广泛的应用。笔者分析了VGI数据应用于地图制图与更新时面临的3大问题,包括数据质量评价、几何数据融合处理和属性数据匹配转换,针对存在的问题提出了相应的解决策略,总结了VGI数据融合处理的整体技术方案和技术流程,并用实例验证方案和流程的可行性。

参考文献:

- [1] 张红平,顾学云,等.志愿者地理信息研究应用进展与趋势[J].地理信息世界,2012,08(04):67-71.
ZHANG Hongping, GU Xueyun et al. Volunteers geographic information research progress and trend[J]. Journal of Geographical Information World,2012,08(04):67-71.(in Chinese)
- [2] Goodchild M F. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0[J]. International Journal of Spatial Data Infrastructures reearch,2013,06(2):23-25.
- [3] 李德仁,钱新林.浅论自发地理信息的数据管理[J].武汉大学学报(信息科学版),2010,35(4):379-383.
LI Deren, QIAN Xinlin. Shallow theory of spontaneous geographic information data management[J]. Journal of Wuhan University(Information Science Edition),2010,35(4):379-383.(in Chinese)
- [4] 单杰,秦昆等.众源地理数据处理与分析方法探讨[J].武汉大学学报(信息科学版),2014,39(4):390-394.
SHAN Jie, QIN Kun, et al.The source of geographical data processing and analysis method to explore[J]. Journal of Wuhan University(Information Science Edition),2014,39(4):390-394.(in Chinese)
- [5] Goodchild M F. Commentary: whither VGI? [J]. GeoJournal,2008,72(4):239-244.
- [6] Jonathan Bennett. OpenStreetMap: Be your own Cartographer[M]. UK: Packt Publishing,2010.
- [7] 钱新林.面向自发地理信息的空间数据表达与管理方法研究[D].武汉:武汉大学,2011.
QIAN XINLIN. Spontaneous expression and management of spatial data of geographic information oriented method research[D]. Wuhan: Wuhan University,2011.(in Chinese)
- [8] Díaz L, Granell C, Gould M, et al. Managing user-generated information in geospatial cyberinfra-structures[J]. Future Generation Computer Systems,2011,27(3):304-314.
- [9] Cooper A K, Rapant P, Hjelmager J, et al. Extending the formal model of a spatial data infrastructure to include volunteered geographical information[C]//25th International Cartographic Conference(ICC 2011).[S.L.]:IEEE,2011,7:4-8.
- [10] OpenStreetMap Data in Layered GIS Format: Version 0.6[EB/OL]: Open Street Map.org,2014.07.25.
- [11] 曾衍伟,龚健雅.空间数据质量控制与评价方法及实现技术[J].武汉大学学报(信息科学版),2004,29(8):686-689.
ZENG Xianwei, GONG Jianya. Spatial data quality control and evaluation methods and implementation technology[J]. Journal of Wuhan University(Information Science Edition),2004,29(8):686-689.(in Chinese)
- [12] 朱庆,陈松林.关于空间数据质量标准的若干问题[J].武汉大学学报(信息科学版),2004,29(10):863-867.
ZHU Qing, CHEN Songlin. Several problems about spatial data quality standard[J]. Journal of Wuhan University (Information Science Edition),2004,29(10):863-867.(in Chinese)
- [13] 刘若梅,周旭.地理信息数据质量的概念、评价和表述—地理信息数据质量控制国家标准核心内容浅析[J].地理信息世界,2008,6(2).
LIU Ruomei, ZHOU Xu. Geographic information data quality concept, evaluation and expression of-national geographic information data quality control standard core content analyses[J]. Journal of Geographical Information in the World, 2008,6(2).(in Chinese)