

面向众源开放街道地图空间数据的质量评价方法

王 明¹ 李清泉^{2,3} 胡庆武¹ 周 檬¹

(1 武汉大学遥感信息工程学院,武汉市珞喻路129号,430079)

(2 武汉大学测绘遥感信息工程国家重点实验室,武汉市珞喻路129号,430079)

(3 深圳大学深圳市空间信息智能感知与服务重点实验室,深圳市南海大道3688号,518060)

摘 要:为了解决来自非专业用户协同编辑的众源地理空间数据质量问题,提出了一种众源地理空间数据质量评价技术框架。针对众源开放街道地图(open street map, OSM)数据特点,选取数据完整性、属性信息准确性和定位精度三方面质量要素,提出了面向众源开放街道地图(OSM)空间数据的质量模型和计算方法。

关键词:众源地理空间数据;OSM;路网;质量元素;质量评价

中图法分类号:P208

众源地理空间数据是指由大众采集并向大众提供的开放地理空间数据^[1]。与传统地理信息采集和更新方式相比,来自非专业大众的众源空间数据具有数据量大、现势性好、信息丰富、成本低等特点和优势,成为近年来国际地理信息科学领域的研究热点^[2-3]。

在探讨众源地理空间数据的处理方法和应用模式时,对众源地理空间数据进行质量分析是需要研究的首要问题^[4]。由于众源地理空间数据存在信息冗余,缺乏质量信息或质量信息不精确等问题,在应用众源地理空间数据前必须对其建立质量分析模型、评价体系和评价方法。

目前,国外学者已经针对欧洲地区的开放街道地图(open street map, OSM)数据质量问题进行了一些研究,建立了初步的 OSM 质量评估模型^[5-6]。为解决来自非专业用户协同编辑的众源空间数据质量问题,本文提出了一种众源地理空间数据质量评价技术框架,针对众源 OSM 的数据特点,选取数据完整性、属性信息准确性和定位精度三方面质量要素,提出了面向众源 OSM 空间数据的质量模型和计算方法。以 2011 版导航地图为参考,对武汉市的众源 OSM 数据质量进行了分析和评价。

1 OSM 数据质量分析

影响 OSM 数据质量的因素主要分为三个方面:首先,数据的采集和地图的绘制是由缺乏足够地理信息知识和有效培训的非专业人员进行的,其中存在一定的人为误差;其次,采集的数据可能来自不同的数据源,具有不同等级的精度;第三,不同采集者使用的 GPS 不同,不同的 GPS 采集到的数据的精度也存在一定的差异。因此,OSM 数据的精度不能依靠常规的地图精度评定方法来精确评估,简单而有效的方法是选择合适质量要素建立质量评估模型,根据该模型与精度更高的数据进行分析对比来评估其数据质量。

1.1 OSM 数据质量要素选取

OSM 空间数据的质量分析需要建立在统一的空间参考、匹配的地物类型、一致的属性描述基础上,选取合适的质量元素建立质量评估模型,按照模型计算其质量参数,进而通过对各质量参数的综合分析,对 OSM 空间数据的完整性、精度和可用性进行评价。

本文选取数据完整性、属性信息准确性和定位精度三方面质量要素对 OSM 空间数据的质量进行评估。其中,数据完整性由表征几何特性

的长度完整性和表征属性质量的名称完整性两个方面构成,属性信息准确性包括名称准确度和类型准确度两个方面,而定位精度则用来表征数据的几何精度。

1.2 质量要素计算模型

1.2.1 数据完整性

数据完整性包括长度完整性和名称完整性。长度完整性反映道路几何质量和数据覆盖性,名称完整性表征名称属性的完成程度。

长度完整性 Q_L 是 OSM 数据区域覆盖情况的直观体现,反映了 OSM 数据的可用性,可用 OSM 道路数据的总长度 L_{OSM} 占参考地图数据道路总长度 L_R 的比例计算:

$$Q_L = L_{OSM}/L_R \quad (1)$$

道路名称是道路重要属性信息,OSM 道路名称完整性直接反映了 OSM 数据的属性质量。名称完整性包括名称属性完整性和名称长度完整性。名称属性完整性 Q_{S_N} 定义为 OSM 道路数据中命名道路数量 S_{OSM}^N 占参考地图数据命名道路数量 S_R^N 的比例:

$$Q_{S_N} = S_{OSM}^N/S_R^N \quad (2)$$

名称长度完整性以道路长度为度量计算,计算公式为:

$$Q_{S_L} = S_{OSM}^L/S_R^L \quad (3)$$

式中, S_{OSM}^L 为已命名的 OSM 道路长度; S_R^L 为已命名的参考数据道路长度。

1.2.2 数据属性信息准确性

数据属性信息准确性包括数据名称准确度和数据类型准确度,反映了 OSM 数据道路属性信息的准确率。

数据名称准确度 Q_{L_M} 指与参考数据道路名称相符的 OSM 道路长度 L_{OSM}^M 占 OSM 道路长度的比例:

$$Q_{L_M} = L_{OSM}^M/L_{OSM} \quad (4)$$

数据类型准确度用于评价 OSM 数据道路类型的准确度,表示为 OSM 数据道路类型与参考数据类型相符的 OSM 道路长度 L_{OSM}^T 占 OSM 道路长度的比例:

$$Q_{L_T} = L_{OSM}^T/L_{OSM} \quad (5)$$

与数据名称准确度不同,由于不同数据集的道路类型划分规范不同,需要根据数据集元数据建立 OSM 数据与参考数据道路类型间的对应关系。

1.2.3 数据定位精度

定位精度是评价众源地理空间数据几何精确度最重要的指标,也是评价众源地理空间数据可用性的重要指标。对于 OSM 道路数据的定位精

度,本文选取文献[7]提出的评估线状要素定位精度的缓冲区分析方法。该方法对参照道路数据按道路实际宽度建立缓冲区,以落在对应参考道路数据缓冲区中的 OSM 道路比例作为 OSM 数据的定位精度 Q_{L_P} :

$$Q_{L_P} = L_{OSM}^P/L_{OSM} \quad (6)$$

式中, L_{OSM}^P 为落在参考道路数据缓冲区中的 OSM 道路长度。

2 实验与分析

2.1 实验数据

研究区域为湖北省武汉市城区,总面积约为 948.46 km²,包括洪山区、武昌区、青山区、江岸区、江汉区、硚口区、东西湖区、汉阳区等 8 个行政区域共 1 471 条道路,道路总长约为 3 466 km。本文实验 OSM 数据来源于 OpenStreetMap 网站,为 WGS-84 经纬度坐标系;参考数据为北京四维图新生产的 2011 版导航地图,定位精度约为 4 m。

2.2 实验结果分析

武汉地区 OSM 数据精度评定实验结果如表 1 所示。

表 1 武汉地区 OSM 精度评定实验结果

Tab. 1 Results of OSM Accuracy Evaluation in Wuhan

质量要素	数据	
长度完整性 Q_L	L_{OSM}	L_R
	3 465 977.447 m	9 129 950.159 m
名称属性完整性 Q_{S_N}	S_{OSM}^N	S_R^N
	1 471	4 092
名称长度完整性 Q_{S_L}	S_{OSM}^L	S_R^L
	2 380 133.9 m	9 129 950.159 m
名称准确度 Q_{L_M}	L_{OSM}^M	L_{OSM}
	1 780 632.921 m	3 465 977.447 m
类型准确度 Q_{L_T}	L_{OSM}^T	L_{OSM}
	1 116 337.777 m	3 465 977.447 m
定位精度 Q_{L_P}	L_{OSM}^P	L_{OSM}
	1 785 989.309 m	3 465 977.447 m

由表 1 可知,武汉地区 OSM 数据的长度完整性为 38.0%,以名称数量为度量的名称完整性为 36.0%,以长度为度量的名称完整性为 26.1%,名称准确度为 51.4%,类型准确度为 32.2%,定位精度为 51.5%。

2.2.1 长度完整性

由表 1 可以看出,整体而言,OSM 数据中道路数据的长度完整性程度较低。由对比分析可知,道路数据长度完整性比较好的地区都是城市

发展程度比较高、人口比较密集的地区,用户上传的道路数据相对比较丰富;而与城市中心距离较远、发展程度低、人口密度小的地区所采集的数据量相对不足,道路数据完成程度也较低。

为了进一步说明地区间差异与 OSM 道路数据完整程度之间的关系,本文对武汉市各个地区的长度完整性指标进行了统计,结果如图 1 所示。由图 1 中可以看出,道路数据完成程度最高的是武昌区,达到了 42.8%;其次是汉阳区和洪山区,均超过 35%;而青山区、东西湖区和硚口区的长度完整性指标则相对较低,由此可知,众源 OSM 数据完成情况受到由城市发展程度、人口密度等因素的制约。

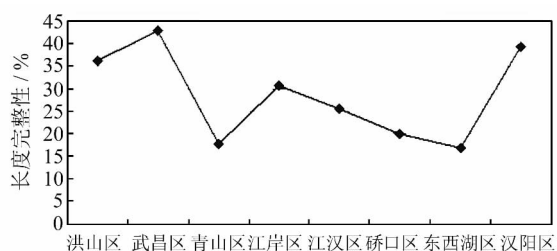


图 1 武汉市各行政区长度完整性

Fig. 1 Length Completeness of District in Wuhan

此外,本文还对各类型道路的长度完整性进行了统计,结果如图 2 所示。可以看出,高等级道路的完成情况较好,比如高速公路(64.2%)、都市高速公路(90.4%)以及国道(72.4%)等,而低等级道路的完整性均不足 15%。最能体现整体长度完整性的道路等级是乡镇村道,完整性为 40.4%,原因是该类型道路占实验数据的比重最大,达到 37.8%。高等级道路数据完成程度相对较好而城市中的小路、人行道等完成程度不佳,也从一定程度上说明众源数据采集工作中普通非驾车用户的参与程度不高,用户提供的众源数据量不足。



图 2 武汉市各类型道路长度完整性

Fig. 2 Length Completeness of Road in Wuhan

2.2.2 名称完整性

由表 1 可知,以道路名称数量为度量的武汉地区 OSM 数据的名称完整性为 36.0%。名称完整性不高的主要原因是 OSM 数据的完成情况即

长度完整性不高(38.0%),从而导致 OSM 中带有名称属性的道路较少。若将名称完整性的计算方法改为 OSM 数据中具有名称属性的道路数量与 OSM 数据道路总数的比值,OSM 数据的名称完整性就能达到 73.3%(见表 2)。这说明在 OSM 数据的已完成道路数据中,其名称完整性较高。另外,OSM 实验数据中未命名的道路主要集中在江夏等城郊地区。导致这种现象的原因除了该地区用户提供的 OSM 数据的数据量有限之外,市政部门对这些偏远地区的道路管理规范程度较城市中心地区不足也是一个原因。

表 2 更改评价方法后的名称完整性结果

Tab. 2 Name Completeness with Improved Method

名称属性完整性	$S_{OSM}^{Q_{S_N}}$	$S_{OSM}^{Q_{S_L}}$	73.3%
Q_{S_N}	1 471	2 006	
名称长度完整性	$S_{OSM}^{L_{OSM}}$	L_{OSM}	68.7%
Q_{S_L}	2 380 133.9 m	3 465 977.447 m	

此外,由表 2 可知,更改评价方法后以道路名字为度量的名称完整性为 73.3%,以道路长度为度量的名称完整性为 68.7%。对比可知,长距离道路的名称完整性比短距离道路的名称完整性低。

2.2.3 名称准确度

由表 1 可知,武汉地区 OSM 数据的名称准确度为 51.4%。影响 OSM 数据名称准确度的因素有很多,主要包括 2011 版导航地图数据中存在未命名的道路数据,OSM 数据中存在无法匹配的道路数据,相连接道路之间的过渡路段归属问题以及立交桥与下方道路在二维数据中的重叠问题等。

为了评定诸如 OSM 多余数据项及导航数据未命名道路数据等因素对名称准确度的影响程度,剔除上述数据项后计算得到的名称准确度结果为:剔除 OSM 多余数据项后的名称准确度为 51.4%,剔除底图未命名道路数据后的名称准确度为 29.1%,两者均剔除后的名称准确度为 58.1%。可以看出,OSM 多余数据项自身对名称准确度影响不大,原因是这部分数据中命名道路与未命名道路所占比例大致相同;而导航数据中的未命名道路数据对结果影响较大,这是因为在剔除这部分底图数据后,许多在原评定方法中判定为命名正确的 OSM 未命名道路数据将不再参与实验,而这部分数据在实验数据中占有相当的比例;两者均剔除后名称准确度指标有显著增长,原因在于该操作在保持命名正确道路数据量的同时,剔除了更多的多余数据,导致比值增加。

此外,各数据集中道路命名的差异对 OSM 数据名称准确度也存在一定的影响。若将存在命名规范问题的道路数据人为提取出来并加以改正,则会对实验结果产生一定的影响。在补充实验中,共提取出总长度为 240 545.62 m 的可直接改正的 OSM 道路数据,将其名称改正后算作名称正确的数据项得到的名称准确度为 58.3%。虽然个人对于名称匹配的判断会带有一定的主观性,但其结果能够定性地说明这一类情况对名称准确度指标造成的影响。

2.2.4 类型准确度

由表 1 可知,武汉地区 OSM 数据的类型准确率为 32.2%,这一指标相对较低。对于整个武汉地区而言,各行政区道路的类型精度相差不大,影响 OSM 数据类型准确度的主要原因是 OSM 数据中存在很多类型属性不正确的长距离主干道路。造成 OSM 数据中主干道路等级属性不正确的主要原因是 OSM 数据的道路分类标准和 2011 版导航地图数据道路分类标准不一致,从而导致两数据集之间的道路类型匹配比较困难。其中最显著的差别是 OSM 道路类型中不存在“国道”,因而 OSM 中很多与 2011 版导航地图数据中类型为“国道”的道路数据对应的道路数据均被判定为类型错误的数据。而 OSM 数据中与 2011 版导航地图中“国道”类型道路相对应的道路数据总长度为 205 040.300 m,占 OSM 数据道路总长度的比值为 5.9%,因而对类型准确度实验的结果产生了较大的影响。

2.2.5 定位精度

按照定位精度评价模型计算,OSM 数据总体定位精度为 51.5%。而在剔除 OSM 中的多余的数据记录后,OSM 数据的定位精度为 60.4%。由此可见,OSM 数据中的多余数据记录对 OSM 数据的定位精度具有较大影响。对 OSM 数据中不同等级道路的定位精度评价结果如图 3 所示。

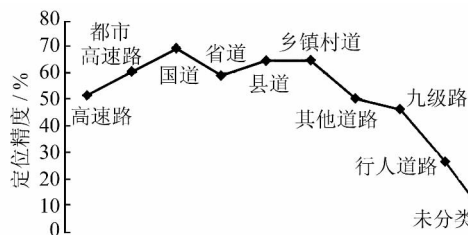


图 3 不同等级道路定位精度

Fig. 3 Position Accuracy of Different Rank Road

从图 3 中可以看出,总体而言,高等级道路比低等级道路定位精度更高。导致低等级道路中行

人道路定位精度特别低(26.6%)的主要原因是人行道上行人很少有意识地进行道路数据的搜集和上传,从而导致用户上传数据量不足。另外,武汉城区道路建设频繁,行人道路边界范围不易界定也有一定的影响。

此外,本文针对武汉市各行政区的定位精度进行了实验,其结果如表 3 所示。从表 3 中可以看出,东西湖区、硚口区等较偏远但精度较高的地区的道路组成以国道、乡镇村道等因等级较高或者数据量较大而导致精度较高的道路类型为主,低精度的行人道路、九级路以及未分类道路的比例较其他区域更少。根据这一现象,在表 5 中添加了反映高精度与低精度道路类型比例的属性列,从中可以看出行政区的道路定位精度与其道路组成中高、低精度道路类型比例之差基本上是正增长的关系。这一现象说明武汉城区中心地区道路数据完成度较高但定位精度与较偏远地区相比并没有明显提升的主要原因是城市中心地区低等级道路较多,城市建设活动频繁。

表 3 武汉各行政区 OSM 数据定位精度/%

Tab. 3 OSM Position Accuracy of District in Wuhan/%

区域名称	高精度道路比例	低精度道路比例	两者差值	定位精度
青山区	30.8	40.8	-10.0	42.1
洪山区	46.3	17.3	29.0	57.6
汉阳区	59.3	17.6	41.7	57.8
江岸区	71.4	13.9	57.5	58.4
武昌区	64.1	12.9	51.2	59.6
东西湖区	50.4	1.3	49.1	68.8
硚口区	68.0	2.6	65.4	71.4
江汉区	88.9	0.8	88.1	78.6

3 结 语

众源地理空间数据的出现,为地理空间数据更新提供了丰富的数据源,但众源数据的质量问题直接影响着众源数据的使用。本文以武汉市为例,对中国地区 OSM 数据的质量进行了评估,并分析了影响质量的因素。评估结果显示,武汉地区 OSM 数据的完整性和定位精度都较低,但就分类道路而言,都市高速路和国道的完整性和定位精度都较高。除此之外,OSM 数据在武汉市部分区域比 2011 版导航地图数据更为详尽。因此,OSM 数据可以成为城市高等级交通基础路网和中心城区道路网数据获取和更新的一种新的来源途径,补充和局部替代传统城市道路网数据采集和更新方法。

参 考 文 献

- [1] 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): 379-382
- [2] Heipke C. Crowdsourcing Geospatial Data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65:550-557
- [3] Goodchild M F. Commentary: Whither VGI? [J]. GeoJournal, 2008, 72: 239-244
- [4] Goodchild M F. Citizens as Sensors: The World of Volunteered Geography[J]. GeoJournal, 2007, 69: 211-221
- [5] Ather A. A Quality Analysis of Open Street Map Data[D]. London: University College, 2009
- [6] De Leeuw J, Said M, Ortega L, et al. An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya[J]. Remote Sensing, 2011, 3(2):247-256
- [7] 李德仁, 王树良, 李德毅. 论空间数据挖掘和知识发现的理论与方法[J]. 武汉大学学报·信息科学版, 2002, 27(3):221-234

第一作者简介: 王明, 博士生, 主要从事众源签到数据分析与挖掘研究。
E-mail: wangming19880910@126.com

Quality Analysis on Crowd Sourcing Geographic Data with Open Street Map Data

WANG Ming¹ LI Qingquan^{2,3} HU Qingwu¹ ZHOU Meng¹

(1 School of Remote Sensing and Information Engineering, Wuhan University,
129 Luoyu Road, Wuhan 430079, China)

(2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3 Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University,
3688 Nanhai Avenue, Shenzhen 518060, China)

Abstract: To solve the quality problem of crowd sourcing geographic data obtained from non-professionals, a quality analysis model for open street map(OSM) crowd sourced geographic data is proposed. Aimed at the data characteristics of OSM, an OSM spatial data quality model with a calculation method is presented with data completeness, attribute accuracy, and positioning accuracy as three quality elements.

Key words: crowd sourcing geographic data; open street map(OSM); road network; quality elements; quality assessment

About the first author: WANG Ming, Ph D candidate, majors in data analysis and mining of crowd sourcing check-in data.
E-mail: wangming19880910@126.com