

国内主流在线地理编码服务质量评价

田沁¹ 巩玥¹ 亢孟军¹ 孟社宁² 杜清运¹

1 武汉大学资源与环境科学学院,湖北 武汉,430079

2 中国船舶重工集团 713 研究所,河南 郑州,450000

摘要:在线地理编码服务是非专业用户最常用的非空间数据空间化的手段。然而,大量在线地理编码服务的出现增加了用户选择的困难,因此通过对主流的在线地理编码服务质量进行评价,为公众的日常应用提供选择上的指导,并有助于在线地理编码服务商发现服务存在的缺陷,进一步完善服务。基于国内主流的 4 种在线地理编码服务——百度、高德、搜狗和腾讯,采用与衣食住行相关的 4 类地址作为检验数据,分别从地址匹配率、地理编码服务位置精度和一致性三个方面,对 4 种服务进行质量评价,得到以下结论:造成 4 种服务质量差异的主要原因是其参考库的质量差异;高德地图有最高的地址匹配率和最差的位置精度;腾讯地图整体表现较为优异,具有较高的数据质量和较完备的地址数据。

关键词:在线地理编码;地址匹配度;质量评价

中图法分类号:P208

文献标志码:A

地理编码是指按照一定的规则赋予个体地域唯一、可识别的编码,建立个体地域与标准地址、空间坐标的映射关系,从而可将地址与空间坐标进行自动转换^[1-4]。地理编码包含 4 个组成部分:输入数据、输出数据、地址匹配算法和参考数据库^[1]。主流的 GIS 软件已普遍包含地理编码模块,如 OneMap、MapInfo 的 MapMarker 等。这些地理编码软件的服务对象主要为企业或者专业用户,需要用户提供参考库,并进行相关过滤参数的设置,普通用户难以使用。

随着互联网地图的迅速发展,在线地理编码服务成为普通用户获取空间位置信息的重要手段。然而,大量在线地理编码服务的出现增加了用户的选择难度,截止 2012 年 9 月,有 162 家单位获得全国甲级互联网地图服务测绘资质,至少有 10 种以上互联网地图及其在线地理编码服务为公众所熟知。因此,正确评价在线地理编码服务的质量,能够为普通用户的日常应用提供指导,并有助于在线地理编码服务商进一步完善相关功能。

地理编码结果的不确定性由地理编码算法、参考库和输入数据三个因素引起^[5]。在使用相同参考库的情况下,不同地理编码算法的计算结果不会对匹配质量产生较大的影响^[2];参考库,即互

联网地图服务的专题要素库(point of interest, POI),其容量、数据质量极大影响了地理编码的结果,是造成不确定性的主要因素^[6];同时,普通用户输入地址的不规范也是造成不确定性的重要因素。

国内学者对于地理编码质量评价的研究相对较少,而主要关注于地名和地址要素的解析方法^[7-10],地理实体及其空间关系的标注^[11-13],试图建立自然语言与 GIS 模型的语义映射。国外学者对地理编码的研究比较全面,涉及到地理编码质量的各个细节,如在公共健康和流行病学研究领域,对一些成熟 GIS 软件中的地理编码模块或商用的地理编码软件包进行地理编码质量的评价^[14-16];在农业健康研究中,比较地理编码质量在农村和城镇区域的差异^[17];利用在线地理编码服务作为农业市场数据验证和精度优化的手段^[18];尝试构建地理编码服务质量评价的系统框架^[19];引入公众参与地理编码结果的评价^[20];甚至引入公众社交行为模式等手段提高地理编码服务质量^[21]。

本文选择百度地图、高德地图、搜狗地图和腾讯地图 4 个主流在线地理编码服务,从地理编码匹配率、位置精度和相似性三个指标进行质量评

价,发现其存在的缺陷,为普通用户选择合适的地理编码服务提供指导。

1 质量评价指标

地理编码质量评价涉及以下4个指标。

(1) 地址匹配度:指原始地址经过地理编码运算的结果与其目标地址的契合程度。常用的计算方法是矢量空间模型(vector space model, VSM),即将原始地址描述和地理编码运算结果看成是一组正交词条矢量 $\{t_1, t_2, \dots, t_m\}$ 所构成的矢量空间,给每一词条 t_i 赋以一定的权重,地址匹配度可以用两矢量的夹角余弦值来度量,值越大,相似度越高^[22]。根据地址匹配度的差别,可将地址匹配度分为精确匹配、模糊匹配和无匹配三种类别计算。

(2) 地址匹配率:指经过地理编码运算,获得匹配结果的地址数占总地址数的比例。匹配率基于一定的匹配度类别。

(3) 位置精度:由位置误差距离反映。位置误差距离是指经过地理编码运算后,获得的匹配结果的坐标与该地址真实坐标的误差距离。误差距离越小,位置精度越高。

(4) 相似性:任意一条地址 ad_i ,经过任意两个在线地理编码服务 s_i 、 s_k 运算,得到两个坐标 $A(x_i, y_i)$ 和 $B(x_k, y_k)$,用 AB 两点间的距离度量服务 s_i 和 s_k 的相似性,距离越短,相似性越高。

2 实验及分析

2.1 数据情况及预处理

选择深圳市与衣、食、住行相关的4类地址数据——购物中心、连锁餐饮、宾馆酒楼和交通设施,共3 959条地址,作为实验样本,数据类别如表1所示。地址组织模式主要有以下两种。

(1) 公共设施名称明确,地址信息表述为:〈地址描述〉=〈行政区划〉+〈设施标准名称〉,如金湖旅馆其地址描述为“深圳市南山区南头街道大新社区”+“金湖旅馆”;

(2) 公共设施名称不明确,地址信息表述为:〈地址描述〉=〈行政区划〉+〈基本区域限定物〉,如书报亭名称可能同一地区出现多次,为不明确名称,其地址描述为“深圳市福田区梅林街道梅亭社区”+“书报亭”。

对4类数据的地址依次调用4个在线地理编码服务接口,并依次记录其返回信息,包括匹配坐标、匹配度及匹配到的地址级别。国家规定公开

出版的电子地图数据必须经过GCJ-02加密,各互联网地图或遵循此规定,或在此基础上进行二次加密,因此匹配后的坐标含有一定的随机误差。为使匹配后的坐标能够进行统一比较,本文选择百度地图提供的百度坐标转换服务接口,将其他4个地理编码服务坐标和测试地址的原始坐标统一转换为BD-09百度坐标(m),在统一的坐标系下进行精度分析。

表1 实验数据类别及数量

Tab. 1 Category and Number of Testing Address

公共设施类别	子类别	子类别数量	小计
购物中心	便利店	199	989
	书报亭	197	
	特色市场	196	
	专卖店	200	
	综合商场超市	197	
连锁餐饮	连锁餐饮	1 000	1 000
馆酒楼	宾馆酒店	500	977
	普通招待所	477	
	公共站点	98	
交通设施	加油站	197	993
	交通运输物流	200	
	汽车维修养护	200	
	汽车售票点	198	
	其他	100	
合计			3 959

2.2 实验结果及分析

2.2.1 地址匹配率

为有效统计各地理编码服务的匹配率,根据各服务的匹配度和返回参数,将地理编码结果分为精确匹配、模糊匹配和未匹配三类。其中,精确匹配是指匹配度高,匹配地址级别达到门牌号、楼栋号或兴趣点的匹配结果;模糊匹配是指匹配度相对较低,匹配地址级别达到道路、社区或者小区级别的一类匹配结果;未匹配是指匹配结果为或匹配地址级别在社区以上的一类匹配结果。基于以上三个匹配级别,统计4类地理编码服务的匹配率,统计结果见表2。

1) 总体分析

对精确匹配、模糊匹配和未匹配分别用0、1和2进行编码,采用Friedman方法对总体匹配率统计结果进行检验。假如每个地址编码服务匹配率都是一样的,则各服务的秩均值应为 $(4+1)/2=2.5$ 。通过该假设检验得到 $p=0.001 \ll 0.05$,其中 p 代表结果的统计学意义,即如果原假设是正确的,那么所得到的样本结果会像实际观测结果那么极端或者更极端的概率为0.001,因此拒绝原假设,即各服务的地址匹配率存在显著差异;然后进一步用

Wilcoxon 符号秩检验对各服务匹配率进行两两配对验证,除了高德和百度的 $p=0.035<0.05$,其他的检验 $p\ll 0.05$,即说明各服务之间确实存在差异

性。根据总体分析的秩均值可知,高德的匹配率最高,其次为腾讯,最后为百度和搜狗。图 1 为各地理编码服务匹配率的秩均值分布。

表 2 地理编码服务的匹配率

Tab. 2 Address Match Rates of Geocoded Results

	数据/条	匹配级别	匹配率/%			
			百度	高德	腾讯	搜狗
所有数据	3 959	精确匹配	39.71	58.22	42.11	37.84
		模糊匹配	39.53	29.33	45.08	45.84
		未匹配	20.76	12.45	12.81	16.31
交通设施	993	精确匹配	42.80	61.53	44.91	43.40
		模糊匹配	34.24	23.26	41.29	39.68
		未匹配	22.96	15.21	13.80	16.92
宾馆酒楼	977	精确匹配	37.26	51.48	40.84	33.88
		模糊匹配	40.22	37.26	44.42	48.72
		未匹配	22.52	11.26	14.74	17.4
购物中心	989	精确匹配	38.32	63.60	36.50	36.30
		模糊匹配	39.74	22.95	51.16	48.43
		未匹配	31.94	13.45	12.33	15.27
连锁餐饮	1 000	精确匹配	40.4	56.2	46.1	37.7
		模糊匹配	43.9	33.9	43.5	46.6
		未匹配	15.7	9.9	10.4	15.7

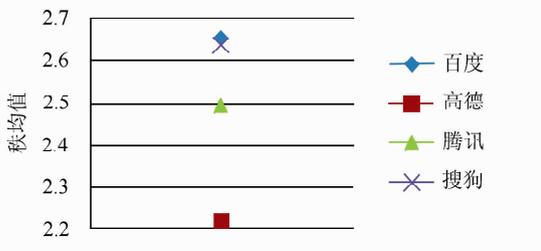


图 1 4 种服务地址匹配率的 Friedman 检验的秩均值
Fig. 1 Mean Ranks of Four Geocoding Services Based on All Data

2) 分类分析

按类别对衣、食、住、行 4 类地址数据的匹配结果的匹配率进行 Friedman 检验,结果发现,各服务的地址匹配率仍然存在显著的差异 ($p\ll 0.005$)。为明确两两服务间的差异值,对 4 种类别地址匹配率数据进行 Wilcoxon 符号秩检验,可得到以下结论(如图 2)。

(1) 除百度和搜狗外,其他服务之间的存在显著差异(即 p 值均远小于 0.05)。

(2) 百度和搜狗两个服务,除了在交通设施类别存在显著差异外 ($p=0.007<0.05$),在其他类别中,并无显著的差异(宾馆酒楼、购物中心和连锁餐饮的 p 值分别为:0.471,0.091,0.274)。

(3) 高德在 4 种类别中匹配率均最高,尤其对于购物中心类别的匹配率最高,其次为腾讯,最后为百度和搜狗。

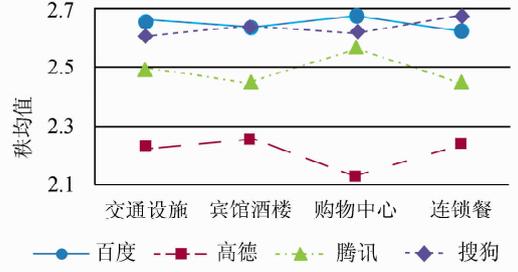


图 2 各种服务对 4 种类别数据的秩均值
Fig. 2 Mean Ranks of Four Geocoding Services Based on Classified Data Types

2.2.2 位置精度

位置精度分析包含总体分析和分类分析。其中,总体分析是指不考虑地址的类别,对所有地址的位置误差距离进行整体的统计,并检验各地理编码服务地址匹配的位置精度是否存在显著差异;分类分析是根据地址的类别(交通设施、宾馆酒楼、购物中心、连锁餐饮),分析不同地理编码服务对各类别地址的匹配位置精度差异。

1) 总体分析

图 3 为 4 种地理编码服务的位置误差距离统计分布。位置误差距离的统计直方图显示各组数据分布不对称,根据切比雪夫不等式,至少有 $1 - \frac{1}{k^2}$ 的数据落在样本平均值的 k 倍标准差范围内。本文采用 4 倍标准差阈值去掉异常值,即至少有

94%的数据落在平均数加减4个标准差的范围之内。表3展示了去掉异常值后各地理编码服务的位置误差距离的统计信息,其中,腾讯和高德的中值、平均值和标准差都明显比搜狗和高德的小。

位置误差距离统计数据的分布整体趋于左偏,并且不满足齐方差性,通过W检验(Shapiro, Wilk test)得 $p < 0.05$,正态性假设被拒绝,即不符合正态分布。由于少量地址经过某种服务匹配产生异常值,去掉异常值后,地址误差距离失去配对特征,因此采用K-W检验(Kruskal Wallis test)判断4种地理编码服务位置精度是否存在

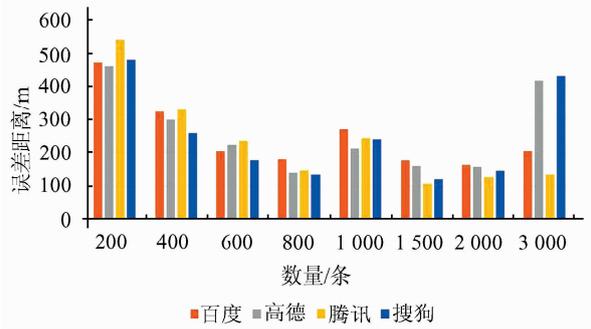


图3 4类地址匹配服务的位置误差距离统计
Fig. 3 Statistics of Location Error of the Four Geocoding Services

表3 地理编码服务匹配结果位置误差统计(去掉异常值)/m

Tab. 3 Statistics of Location Error Without Outliers/m

服务	数目	最小值	最大值	中值	平均值	标准差
百度	3 943	0.61	36 914.42	206.08	765.25	1 701.69
高德	3 942	0.22	44 763.47	228.17	1 227.91	3 392.75
腾讯	3 943	0.76	37 260.31	176.84	587.13	1 534.81
搜狗	3 942	0.53	60 591.40	204.72	1 306.68	3 742.13

显著差异。

K-W检验是一种针对多组数据的非参数检验,用来检测总体函数分布的一致性原假设和其替代假设,判断多组数据间是否存在显著差异。由K-W检验结果可知 $p \ll 0.05$,拒绝原假设,即4个地理编码服务的地址匹配位置精度确实存在显著的差异(如图4);然后两两服务之间分别进行U检验(Mann-Whitney test)来检验这4个服务两两之间是否都存在差异,结果显示百度-高德($p = 0.03$)、百度-腾讯($p < 0.001$)、高德-腾讯($p < 0.001$)以及腾讯-搜狗($p < 0.001$)存在显著性差异,而百度-搜狗($p = 0.135$)和高德-搜狗($p = 0.196$)之间的差异则不明显;秩均值反映了地理编码服务位置精度的高低,值越小,位置误差距离越小,位置精度越高。腾讯的位置精度最高,其次为百度、搜狗和高德。

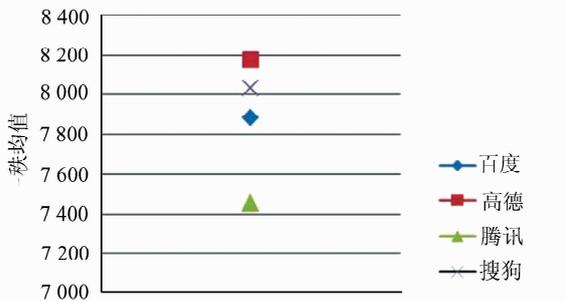


图4 所有地址K-W检验秩均值

Fig. 4 K-W Test Mean Ranks Based on All Address Data

2) 分类分析

对4类地址的位置误差距离进行分类分析,确定4种地理编码服务对不同类别地址的匹配质量的优劣。由于4类地址的位置误差距离数据均不满足方差分析的前提条件,本文仍采用K-W检验方法。

表4为购物中心、连锁餐饮、宾馆酒楼和交通设施4类地址的位置误差的K-W检验结果,可得到以下结论。

表4 4类地址数据的位置误差距离K-W检验结果

Tab. 4 K-W Test Results of Location Error of Four Types of Address

地址类别	服务	数目	秩均值	显著性
购物中心	百度	986	1 996.58	0.001
	高德	986	2 101.92	
	腾讯	986	1 784.01	
	搜狗	986	2 007.49	
连锁餐饮	百度	1 000	2 016.33	0.002
	高德	1 000	1 987.56	
	腾讯	1 000	1 902.01	
	搜狗	1 000	2 095.10	
宾馆酒楼	百度	968	1 912.98	0.001
	高德	968	2 062.32	
	腾讯	968	1 832.23	
	搜狗	967	1 936.47	
交通设施	百度	989	1 969.98	0.187
	高德	988	2 034.84	
	腾讯	989	1 923.20	
	搜狗	989	1 984.04	

(1) 对于购物中心类别, K-W 检验显著性值 $p < 0.01$, 拒绝原假设, 即 4 种地理编码服务对该类地址的地址匹配位置精度差异显著; 然后用 U 检验各服务两两之间的差异性, 结果显示百度-搜狗 ($p = 0.839$) 和高德-搜狗 ($p = 0.082$) 之间的差异性并不明显, 其他服务之间差异性比较显著; 结合秩均值可知, 腾讯位置精度最高, 其次为百度、搜狗和高德。

(2) 对于宾馆酒楼类别, K-W 检验显著性值 $p < 0.01$, 拒绝原假设, 即 4 种地理编码服务对该类地址的地址匹配位置精度差异显著; 由 U 检验显示出百度-腾讯 ($p = 0.097$)、百度-搜狗 ($p = 0.621$)、腾讯-搜狗 ($p = 0.067$) 之间差异性不显著, 其他服务之间差异显著; 结合秩均值比较可知, 高德的位置精度最低, 其他几个服务之间位置精度差异不显著。

(3) 对于餐饮连锁类别, K-W 检验显著性值 $p < 0.01$, 拒绝原假设, 4 种服务在该类别的位置精度差异显著; 同时由 U 检验可知, 百度-高德 ($p = 0.620$)、百度-搜狗 ($p = 0.113$) 及高德-腾讯 ($p = 0.111$) 并不显著; 结合秩均值比较可知, 腾讯的位置精度要优于百度和搜狗。

(4) 对于交通设施类别, K-W 检验显著性值 $p > 0.01$, 接受原假设, 即 4 种服务在该类别的位置精度无显著差异。

2.2.3 相似性

相似性分析分为总体分析和分类分析。总体分析是不区分地址数据的类别, 对各地地理编码服务间的相似性进行分析; 分类分析是按照地址类别, 进行服务间的相似性分析。从百度、高德、腾讯和搜狗 4 个地图服务中任选两个服务组合进行相似性分析, 可得到组合数 6, 即共有 6 组相似性值可供比较。

1) 总体分析

根据切比雪夫不等式, 采用 4 倍标准差阈值去掉异常值。由于数据不满足方差齐性, 并且百度、高德、腾讯和搜狗 4 个地图服务两两之间的相似性值构成的 6 个变量失去了配对的特点, 不满足多个变量相关的条件, 因此先对 6 组相似性值的四分位数进行分析, 然后采用 K-W 方法检验对各服务之间相似性的显著性。

图 5 为 6 组相似性值取自然数为底数的对数, 生成的箱线图。通过对比中位数、数据集聚程度, 可以看出百度-腾讯的相似性最大, 高德和搜狗的相似性最小。

对各服务的相似性进行 K-W 检验, 得到 $p <$

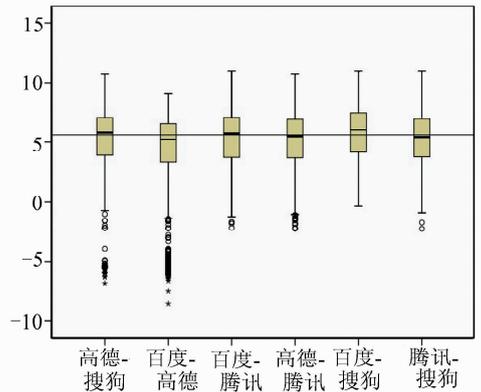


图 5 地理编码服务相似性箱线图
Fig. 5 Boxplot of Similarity Among Four Geocoding Services

0.001, 即在服务之间的相似性上差异性显著的; 然后对这 6 对相似值进行 U 检验, 发现只有百度-高德与百度-搜狗之间的差异是不显著的 ($p = 0.133$), 其他的相似值之间差异是比较显著的; 结合秩均值, 可以得出百度-腾讯最为相似, 其次为高德-腾讯、腾讯-搜狗、(百度-搜狗、百度-高德)、高德-搜狗(如图 6)。

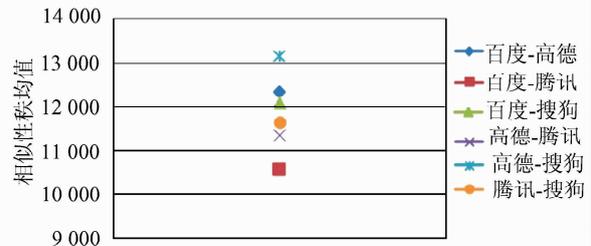


图 6 相似性秩均值对比
Fig. 6 Visualization of K-W Test Mean Rank About Similarity Among Services

2) 分类分析

表 5 对 6 组数据的相似性值用四分位数进行统计分析, 其具体结果说明如下。

(1) 宾馆酒楼类。百度-腾讯的相似性最大, 其后依次为腾讯-搜狗、高德-腾讯、百度-搜狗、百度-高德和高德-搜狗。

(2) 购物中心类。百度-腾讯相似性最大, 其后依次为高德-腾讯、搜狗-腾讯、百度-搜狗、百度-高德和高德-搜狗。

(3) 交通设施类。相似性依次为百度-腾讯、腾讯-搜狗、百度-搜狗、高德-腾讯、百度-高德和高德-搜狗。

(4) 餐饮连锁类。相似性依次为百度-腾讯、高德-腾讯、腾讯-搜狗、百度-高德、百度-搜狗和高德-搜狗。

总结可知,百度-腾讯在所有类别地址中都有最大的相似性,而高德-搜狗的相似性则最小;同

时,连锁餐饮类各个服务之间的相似性最大,其后依次为购物中心,交通设施,宾馆酒楼。

表5 4种地理编码服务相似性的四分位值/m
Tab.5 Quartiles of Similarity Among Four Geocoding Services/m

		百度-高德	百度-腾讯	百度-搜狗	高德-腾讯	高德-搜狗	腾讯-搜狗
所有数据 (去掉异常值)	Q1	52.17	30.40	44.31	37.43	65.50	45.09
	Q2	339.57	188.52	316.38	230.92	432.77	232.94
	Q3	1 238.60	730.58	1 249.22	1 012.95	1 786.87	1 077.88
宾馆 酒楼	Q1	77.69	39.13	49.10	46.94	86.74	42.22
	Q2	450.54	231.93	364.33	289.69	461.67	274.65
	Q3	1 197.89	839.24	1 249.19	1 094.81	1 831.62	1 170.80
购物 中心	Q1	66.31	34.22	57.91	43.57	93.90	52.75
	Q2	327.80	164.52	323.37	166.83	408.45	201.73
	Q3	1 180.75	654.22	1 045.24	804.09	1 651.62	832.15
交通 设施	Q1	39.28	26.00	34.83	31.40	47.95	43.49
	Q2	369.97	198.90	281.87	297.14	459.12	231.85
	Q3	1 530.65	762.06	1 279.99	1 194.90	1 779.68	1 035.82
连锁 餐饮	Q1	43.35	29.58	44.82	26.61	55.99	40.68
	Q2	243.35	153.10	287.91	184.34	391.45	214.26
	Q3	1 110.55	678.51	1 422.34	891.91	1 854.06	1 200.68

3 结 语

地理编码是将非空间数据空间化最高效、最简单的方法,但是地理编码是一个不确定性的过程,其精度主要受匹配算法、参考库和输入数据的影响。公众在使用这些地理编码服务时,希望了解这些地理编码服务的特点,从而选择适合自己的地理编码服务。本文针对这4个服务商的地理编码服务进行了分析,从匹配率、位置精度和相似性3个维度进行总体对比和分类对比,得到以下结论。

1) 地址匹配率。总体上,高德的匹配率最高,其次为腾讯,最后为百度和搜狗(二者差异性不显著)。分类分析中,高德在4类地址的匹配率都是最高的,特别是购物中心的匹配率远高于其他几个服务;其次为腾讯,最后为百度和搜狗,其中,搜狗除了在交通设施类数据中匹配率略高于百度外,在其他类别数据中两者差异均不显著。

2) 位置精度。总体上,4种服务可分成两个级别,腾讯位置精度最高,为第一级别;其他三种服务与腾讯位置精度差异显著,为第二个级别,其顺序为百度、搜狗和高德。分类分析中,4种服务在交通设施类别差异不显著;购物中心类别位置精度顺序跟总体分析结论一致;宾馆酒楼类别,高德的位置精度最低,其他的几个服务差异不显著;连锁餐饮类别,腾讯位置精度要优于百度和搜狗。

3) 相似性。腾讯和百度匹配的结果相似性

最高,其次为高德-腾讯,然后是腾讯-搜狗,最后是百度-搜狗和百度-高德(这两者的相似性差异不显著),高德和搜狗的匹配结果相似性最小;在对数据按类别进行分析时可知,百度-腾讯在所有类别中均相似度最高,而高德-搜狗的相似性最低;并且对于连锁酒店数据,各服务之间的相似性最大,其次为购物中心、交通设施、宾馆酒楼。

综合以上3个指标的结论可知,高德有最高的地址匹配率和最差的位置精度,说明其地址参考库数据相对较为完备,地址匹配算法完善,但其参考地址的数据精度较差;腾讯整体表现较为优异,具有较高的数据质量和较完备的地址数据,而搜狗则为4者中最差的。

在分类分析方面,4种服务在交通设施类的地址的匹配结果较好,而且无显著差异,原因可能与该类地址比较明确,且变更频率低有关;而表现差异显著的主要为连锁酒店类,原因是部分连锁酒店处于小的街巷中,地址描述混乱,难以规范的描述和组织。

以下因素可能对本文结论产生影响:①地址匹配返回的坐标经过了加密,为进行对比分析,统一到BD-09坐标系下,该加密算法对真实坐标进行了非线性偏移;②原始数据由地址描述和其实际坐标组成,原始地址描述在采集过程中虽然遵循了一定的行业推荐标准,但并非强制标准,可能与各服务参考库的地址描述存在一定的差异。因此,需要进一步规范地址模型,提高地理编码服务的质量,促进更多非空间数据的“落地”。

参 考 文 献

- [1] Goldberg D W, Wilson J P, Knoblock C A. From Text to Geographic Coordinates: the Current State of Geocoding [J]. *URISA Journal*, 2007, 19(1): 33-46
- [2] Karimi H A, Durcik M, Rasdorf W. Evaluation of Uncertainties Associated with Geocoding Techniques[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2004, 19(3): 170-185
- [3] Zandbergen P A. A Comparison of Address Point, Parcel and Street Geocoding Techniques [J]. *Computers, Environment and Urban Systems*, 2008, 32(3): 214-232
- [4] Roongpiboonsopit D, Karimi H A. Comparative Evaluation and Analysis of Online Geocoding Services [J]. *International Journal of Geographical Information Science*, 2010, 24(7): 1 081-1 100
- [5] Hart T C, Zandbergen P A. Reference Data and Geocoding Quality: Examining Completeness and Positional Accuracy of Street Geocoded Crime Incidents[J]. *Policing: An International Journal of Police Strategies & Management*, 2013, 36(2): 263-294
- [6] Roongpiboonsopit D, Karimi H A. Quality Assessment of Online Street and Rooftop Geocoding Services[J]. *Cartography and Geographic Information Science*, 2010, 37(4): 301-318
- [7] Zhang Xueying, Lv Guonian, Li Boqiu, et al. Rule-based Approach to Semantic Resolution of Chinese Addresses [J]. *Journal of Geo-Information Science*, 2010, 12(1): 9-16(张雪英, 闰国年, 李伯秋, 等. 基于规则的中文地址要素解析方法[J]. 地球信息科学学报, 2010, 12(1): 9-16)
- [8] Cheng Gang, Lu Xiaoping. Matching Algorithm for Chinese Place Names by Similarity in Consideration of Semantics of General Names for Places[J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(4): 404-410(程钢, 卢小平. 顾及通名语义的汉语地名相似度匹配算法[J]. 测绘学报, 2014, 43(4): 404-410)
- [9] Zhang Xueying, Krause J. An Approach to Automatic Keyword Extraction in Chinese Text [J]. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(4): 512-520(张雪英, Krause J. 中文文本关键词自动抽取方法研究[J]. 情报学报, 2008, 27(4): 512-520)
- [10] Tang Xuri, Chen Xiaohe, Zhang Xueying. Research on Toponym Resolution in Chinese Text[J]. *Geomatics and Information Science of Wuhan University*, 2010, 35(8): 930-935(唐旭日, 陈小荷, 张雪英. 中文文本的地名解析方法研究 [J]. 武汉大学学报·信息科学版, 2010, 35(8): 930-935)
- [11] Zhang Xueying, Zhang Chunju, Du Chaoli. Semantic Relation Between Spatial Relation Terms and Feature Type of Geographical Entities[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(11): 1 266-1 270(张雪英, 张春菊, 杜超利. 空间关系词汇与地理实体要素类型的语义约束关系构建方法[J]. 武汉大学学报·信息科学版, 2012, 37(11): 1 266-1 270)
- [12] Zhang Xueying, Zhu Shaonan, Zhang Chunju. Annotation of Geographical Named Entities in Chinese Text[J]. *Acta Geodaetica et Cartographica Sinica*, 2012, 41(1): 115-120(张雪英, 朱少楠, 张春菊. 中文文本的地理命名实体标注[J]. 测绘学报, 2012, 41(1): 115-120)
- [13] Zhang Xueying, Zhang Chunju, Zhu Shaonan. Annotation for Geographical Spatial Relations in Chinese Text [J]. *Acta Geodaetica et Cartographica Sinica*, 2012, 41(3): 468-474(张雪英, 张春菊, 朱少楠. 中文文本的地理空间关系标注[J]. 测绘学报, 2012, 41(3): 468-474)
- [14] Whitsel E A, Rose K M, Wood J L, et al. Accuracy and Repeatability of Commercial Geocoding[J]. *American Journal of Epidemiology*, 2004, 160(10): 1 023-1 029
- [15] Whitsel E A, Quibrera P M, Smith R L, et al. Accuracy of Commercial Geocoding: Assessment and Implications[J]. *Epidemiologic Perspectives & Innovations*, 2006, 3(1): 1-12
- [16] Zandbergen P A, Green J W. Error and Bias in Determining Exposure Potential of Children at School Locations Using Proximity-Based GIS Techniques [J]. *Environmental Health Perspectives*, 2007, 115(10): 1 363-1 370
- [17] Jones R R, DellaValle C T, Flory A R, et al. Accuracy of Residential Geocoding in the Agricultural Health Study[J]. *International Journal of Health Geographics*, 2014, 13(1): 1-9
- [18] Cui Y. A Systematic Approach to Evaluate and Validate the Spatial Accuracy of Farmers Market Locations Using Multi-geocoding Services[J]. *Applied Geography*, 2013, 41(4): 87-95
- [19] Goldberg D W, Ballard M, Boyd J H, et al. An Evaluation Framework for Comparing Geocoding Systems[J]. *International Journal of Health Geographics*, 2013, 12(1): 1-15
- [20] Sierra J P G-B, Callejo M A M, Garrido R A. Volunteering Assistance to Online Geocoding Services Through a Distributed Knowledge Solution [C].

The RICH-VGI Workshop at 18th AGILE Conference on Geographic Information Science, Lisbon, Portugal, 2015

- [21] McKenzie G, Janowicz K. Where is also about Time: A Location-Distortion Model to Improve Reverse Geocoding Using Behavior-Driven Temporal Semantic Signatures[J]. *Computers, Environment*

and Urban Systems, 2015, 54: 1-13

- [22] Diao Lili, Wang Likun, Lu Yuchang, et al. Computing Similarity Threshold for Text Classification [J]. *Journal of Tsinghua University (Science and Technology)*, 2003, 43(1): 108-111(刁力力, 王丽坤, 陆玉昌, 等. 计算文本相似度阈值的方法[J]. *清华大学学报(自然科学版)*, 2003, 43(1): 108-111)

A Comparative Evaluation of Online Geocoding Services in China

TIAN Qin¹ GONG Yue¹ KANG Mengjun¹ MENG Shening² DU Qingyun¹

¹ School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

² Seven-Thirteen Institute of China Shipbuilding Industry Corporation, Zhengzhou 450000, China

Abstract: Online geocoding services are the most common technique for transforming non-spatial into spatial information. The increasing number of online geocoding services, however make it difficult for requestors to choose a better service. There are two aims in this paper: (1) to provide guidance for service requestors when selecting the most suitable service; (2) to discover the defects of online geocoding services, thus providing a basis for further improvements. In this paper, a comparative evaluation of geocoding quality was conducted among the most four popular online geocoding services: Baidu, Amap, Sougou, and Tencent. Four types of address data associated with basic public necessities were used as test data, and three metrics for geocoding quality: match rate, positional accuracy and similarity, were calculated to evaluate the quality of these four online geocoding services. The following conclusions were drawn: (1) online geocoding service quality mainly depends on the quality of the reference database; (2) Amap produced the highest match rate and lowest positional accuracy; (3) Overall, the Tencent map service performs the best, and produced more complete address data with higher data quality.

Key words: online geocoding service; address matching degree; quality evaluation

First author: TIAN Qin, PhD candidate, specializes in the geocoding and temporal/spatial data mining. E-mail: tianqin@whu.edu.cn

Corresponding author: KANG Mengjun, PhD, lecturer. E-mail: mengjunc@whu.edu.cn

Foundation support: The National Natural Science Foundation of China, Nos. 41201403, 41471327.