

ボランティアな地理空間情報の品質評価指標としての信頼性に関する一考察

山下 潤*

A note on trust as a quality assessment indicator for VGI

Jun YAMASHITA

Abstract: In this note, the quality of volunteered geospatial information was evaluated using the trust, which is a quality assessment indicator other than the data quality elements of geospatial information defined by the ISO. OpenStreetMap (OSM) data were used to assess their data quality. As a result, it was shown that the trust could become an indicator to assess the data quality of VGI.

Keywords: ボランティアな地理情報 (volunteered geographic information, VGI), データ品質評価 (data quality assessment), 信頼性 (trust)

1. はじめに

近年, 日本においても OpenStreetMap (OSM) で代表されるボランティアな地理空間情報 (volunteered geographic information, VGI) に関する学術的研究が蓄積されつつあり, その一分野が VGI のデータ品質評価に関する研究である. この分野で研究が進展した背景として, VGI の作成主体 (先述した OSM の場合は OSM マッパー, OSM contributor) が主に一般市民であることから, 彼らが作成したデータの品質が必ずしも保証されるものではないことがあげられる.

VGI のデータ品質を評価することで, VGI のデータ品質を担保するため, ISO 19157:2013 等のデータ品質要素が評価指標として VGI のデータ品質評価研究で利用されてきた. これらデータ品質要素は完全性, 論理一貫性, 位置正確度, 時間正確度, 主題正確度等で構成される. これらのデータ品質要素を評価指標として, 主に OSM データを用いた VGI の品質評価に関する研究が蓄積されつつある. これらの研究のなかで, Haklay (2010) を嚆矢として, 特に位置正確度に関する研究は国内外を問わず多い (金杉ほか, 2019).

近年 VGI のデータ品質評価研究では, ISO19157:2013 以外の品質要素を用いて, VGI のデータ品質を評価しようとする試みがみられる. Senaratne et al. (2017) は, ISO19157:2013 以外の品質要素として, credibility

(信憑性), reputation (評判), trust (worthiness) (信頼性・信頼) 等があることを指摘した. De Longueville et al. (2010) は credibility を, 意図的もしくは意図しない誤り, 省略, または誇張を考慮しつつも, その情報の一部を信じることができるか否かを説明するための主観的な概念と定義した (p. 43). この定義を踏まえて, Senaratne et al. (2017) は各種の VGI の credibility のうち, 画像型 VGI (主に Flickr 画像) とテキスト型 VGI (主にジオタグ付き Twitter メッセージ) の credibility を中心に従来の研究をある程度展望している. さらに Flanagan and Metzger (2008) は credibility が trust (=worthiness) と expertise (専門性) で構成されていると考えた (p. 141).

このように trust は credibility よりも限定的な概念であるが, Bishr and Kuhn (2007) によってデータ品質の代替尺度として trust が提案されて以降, trust に関する研究が深化されつつある. Keßler and De Groot (2013) は, Sztompka (1999) による定義 (p. 25) を踏まえて, trust を他人の将来的な偶発的行動に関する賭け (bet) と定義する一方で, trust と類似した概念である reputation を, Mezzetti (2004) による定義を踏まえて, ある人物や物事の歴史的行動に関する情報から推測される信頼 (trustworthiness) の主観的な認識と定義した (pp. 23-24). しかし, このようなデ

* 正会員 九州大学大学院比較社会文化研究院 (Kyushu University)
〒819-0395 福岡市西区元岡 744 Tel:092-802-5640

ータ品質評価に関する定性的な品質要素による実証研究は十分に蓄積されておらず、今後当該分野における研究をさらに深化させる必要がある。

一方、データ品質要素以外から VGI のデータ品質の保証に接近しようとした研究として、品質保証へのアプローチの仕方に着目する研究があげられる。Haklay et al. (2010) は、リーナスの法則に基づき、クラウドソーシングによる VGI の品質保証を、OSM の道路データを用いて検討した。Goodchild and Li (2012) は VGI の品質保証に関する 3 つのアプローチを示したが、このうち、ある個人の投稿者による誤りを集団で検証・訂正する方法がクラウドソーシング・アプローチである。Haklay et al. (2010) は、結果として、OSM マッパーの数と位置正確性の関係は統計的に有意でないことを示すことで、クラウドソーシングによる VGI のデータ品質保証に疑義を呈した。同様に、Yamashita et al. (2019) は、クラウドソーシングによる品質保証に関して、OSM 利用者とともに OSM マッパーがデータ品質要素を意識していないことから、クラウドソーシング以外のアプローチならびに上述した 5 つのデータ品質要素以外の要素での VGI の品質評価の必要性を指摘した。

Goodchild and Li (2012) が示したクラウドソーシング・アプローチ以外のアプローチの一つがソーシャル・アプローチである。彼らは、当該アプローチを、VGI への貢献という点で高く評価され、信頼できる個人が、他の投稿者による VGI のゲートキーパーの役割を果たし、VGI の品質を維持・制御することと定義した。この点を踏まえて、Morero et al. (2018) はソーシャル・アプローチにもとづき、trust と reputation を組み込んだ TR モデルを用いて、土地区画に関連する広範な地理空間データの処理が適正に行われているかを分析し、当該モデルの有用性を検証した。

このように、VGI のゲートキーパーを通じて、VGI のデータ品質を担保できるという長所がソーシャル・アプローチにはあるが、VGI 単独で、信頼性 (trust) を評価することは困難である。これに対して、信頼性指標を用いた場合、VGI 単独でそのデータ品質を保証できるという長所がある。しかし次章で詳述する Keßler and De Groot (2013) が示した信頼性指標の有

効性を検証した研究は限られており、当該分野での研究の蓄積は多くない。この点に鑑み、本稿では、まず公共データと VGI を比較し、対象とする VGI で ISO 上のデータ品質が担保されていることを確認した上で、VGI のデータ品質保証に関する上述した 5 つの信頼性指標の有用性を明らかにするを目的とする。この目的を明らかにするため、次章で研究方法を示す。続く 3 章で、研究結果を示し、4 章で、この結果を検討するとともに、今後の課題についても述べる。

2. 研究方法

VGI の信頼性に関しては、Keßler and de Groot (2013) が提示した 5 指標 (バージョン数、ユーザー数、編集確認回数、タグ修正回数、ロールバック数) を用いた (表 1)。バージョン数は、地物が改訂された回数で表される。彼らは、バージョン数が大きいほど、地物の信頼性が高まると考えた。ユーザー数は、地物の作成に関与したユーザーの数で表される。Haklay et al. (2010) が指摘したように、リーナスの法則が働き、ユーザー数が多いほど、信頼性は向上する。編集確認回数は、当該地物の最後の編集以降に、当該地物から半径 50m 以内でのすべての地物の編集回数で表される。編集確認回数が多いと、信頼性も高まると考えられた。タグ修正回数は、地物のタグが変更された回数で定義される。この回数が多い場合、地物の分類があいまいであることを表し、信頼性を低下させると考えられた。ロールバック数は、タグ修正の一種であり、地物のタグを以前のタグに戻した場合、1 回と数えられる。タグ修正回数と同様に、ロールバック数も地物分類のあいまいさを表し、この回数が多いと、信頼性が低くなると考えられた。

日本国内の OSM の山の名称を研究対象として、これ

表 1 分析で用いた信頼度指標

バージョン数
ユーザー数
編集確認回数
タグ修正回数
ロールバック数

ら5指標の値を計算した。OSMの山の名称とは、OSMデータのなかで、タグがnatural=peakとなるノードデータの名称(NAME=*)を指す(以下OSM-NMデータと略)。これらのノードデータに対して、上記の5指標の値を算出するにあたり、OSMのヒストリー ダンプデータを用いた。

加えて、OSMの信頼性指標の算出に加えて、OSMの位置正確度と主題正確度も吟味した。その際、国土地理院ベクトルタイル提供実験で提供されている「地名情報(自然地名)」の「山」ノードデータの「名称」(以下公的NMデータと略)を参照データとして用いた。

上記のデータに対して、以下の2つの分析を行った。第1に、公的NMデータと比較し、OSM-NMデータの位置・主題正確度を調べた。位置正確度に関しては、OSM-NMデータと対応する公的NMデータを抽出し、両者の直線距離を求めた。距離が短いほど、OSM-NMデータの位置正確度は高いと判断した。一方主題正確度に関しては、OSM-NMデータと対応する公的NMデータの名称を比較し、一致もしくは不一致であるかを調べ、一致する場合、主題正確度が高いと判断した。

第2に、公的NMデータとの関係で、上記5指標と信頼性との関係を明らかにした。その際、位置正確度と主題正確度に二分し、重回帰分析を用いて、5指標の影響を吟味した。位置正確度に関しては、上述したOSM-NMデータと公的NMデータの距離を従属変数とする一方で、主題正確度に関しては、公的NMデータの山の名称と一致・不一致した場合、それぞれ1と0で表すダミー変数を従属変数とし、5信頼性指標を独立変数として、重回帰分析を行った。

3. 研究結果

3.1. OSMの位置・主題正確度

結果として、OSMは位置・主題正確度の双方で高い値を示したことから、ISO上のデータ品質が担保されているといえる。まず位置正確度に関して、OSM-NMデータと公的NMデータの平均距離は19.13m、標準偏差は54.34mであった。この平均と標準偏差の値を用いて標準化し、0.5間隔で両データの距離の度数分布表を作成した(表2)。この表でわかるように、±0.5以内に93.6%が分布し、著しく逸脱した値をとっていないこと

表2 位置正確度の評価

階級	度数	構成比(%)
-0.5以上0.0未満	5,929	85.9
0.0以上0.5未満	531	7.7
0.5以上1.0未満	127	1.8
1.0以上1.5未満	55	0.8
1.5以上2.0未満	46	0.7
2.0以上2.5未満	28	0.4
2.5以上3.0未満	28	0.4
3.0以上	159	2.3
合計	6,903	100.0

(平均=19.13m, 標準偏差=54.34m)

がわかる。さらにOSM-NMデータの約86%は、公的NMデータとの距離が平均値より短いことから、公的NMデータに極めて近い場所に位置づけられていることもわかる。このことから、OSM-NMデータの位置正確度は高いと判断した。一方主題正確度に関しては、OSM-NMデータと公的NMデータが一致するケースが全体の95.3%であり、この点から、OSM-NMデータの主題正確度は高いと判断できる(表3)。

3.2. 信頼性指標の有効性

位置・主題正確度と5指標の関係を、位置・主題正確度それぞれに、全6,903データを対象としたケース、一致したケース、不一致したケースの3ケース、計6ケースで検討した(表4)。これらのケースの結果から、5指標が、VGIのデータ品質を保証するための信頼性を表す指標であるとはいいいがたい。

決定係数は全6ケースとも1%水準で統計的に有意であることから、5指標による説明率は高いといえる。特に主題正確度に関しては、従属変数がダミー変数で

表3 主題正確度の評価

	度数	構成比(%)
一致	6,576	95.3
不一致	327	4.7
合計	6,903	100.0

あったこともあり、説明率は非常に高い。

これら5指標に関して、位置正確度は、主にバージョン数とタグ修正回数によって、一方主題正確度は、編集確認回数を除く他の4指標に規定されたと考えられる。これらの指標以外では、位置正確度に関する一致したケースでユーザー数とタグ修正回数が、主題正確度に関する全データを対象としたケースで編集確認回数が、少なくとも10%水準で統計的に有意であり、これらの指標も位置・主題正確度に影響を与えたと考えられる。

しかし位置・主題正確度への5指標の影響は、Keßler and de Groot (2013) の仮説とは異なるものもあった。

まず位置正確度に関して、バージョン数とタグ修正回数は少なくとも10%水準で統計的に有意であった。しかし彼らの仮説にもとづけば、バージョン数は位置正確度を表す変数として使った OSM-NM データと公的 NM データの距離とは負の相関を、逆にタグ修正回数は正の相関を示すべきであるが、重回帰分析の結果から、前者の符号は正であり、後者は負であり、彼らの仮説に反する。唯一、一致のケースのロールバック数のみ、

表4 信頼性指標の位置・主題正確度への影響

a)位置正確度		b)主題正確度	
1)全体		1)全体	
偏回帰係数		偏回帰係数	
バージョン数	20.595 ***	バージョン数	0.877 ***
ユーザー数	-2.730	ユーザー数	0.083 ***
編集確認回数	2.746	編集確認回数	-0.014 *
タグ修正回数	-5.051 *	タグ修正回数	-1.017 ***
ロールバック数	-1.503	ロールバック数	0.496 ***
調整済み決定係数	0.115 ***	調整済み決定係数	0.953 ***
2)一致		2)一致	
偏回帰係数		偏回帰係数	
バージョン数	6.908 ***	バージョン数	0.990 ***
ユーザー数	1.232 *	ユーザー数	0.009 ***
編集確認回数	-0.045	編集確認回数	0.001
タグ修正回数	-6.036 ***	タグ修正回数	-0.991 ***
ロールバック数	2.545 ***	ロールバック数	0.475 ***
調整済み決定係数	0.555 ***	調整済み決定係数	0.997 ***
3)不一致		3)不一致	
偏回帰係数		偏回帰係数	
バージョン数	194.950 ***	バージョン数	1.816 ***
ユーザー数	-19.350	ユーザー数	0.168 ***
編集確認回数	12.010	編集確認回数	0.011
タグ修正回数	-170.980 ***	タグ修正回数	-1.882 ***
ロールバック数	62.160	ロールバック数	1.096 ***
調整済み決定係数	0.628 ***	調整済み決定係数	0.996 ***

Signif. : *** < 0.01, ** < 0.05, * < 0.1.

符号が正であり、かつ1%水準で統計的に有意であることから、彼らの仮説と合致する。以上から、位置正確度に関しては、5指標は信頼性を表す指標であるとはいいたい。

他方主題正確度に関して、編集確認回数を除く、バージョン数、ユーザー数、タグ修正回数、ロールバック数が、3ケースすべてで1%水準で統計的に有意であった。しかしKeßler and de Groot (2013)の仮説にもとづけば、公的NMデータとの一致・不一致をそれぞれ1と0で表したことから、ロールバック数は負の相関を示すべきである。しかし重回帰分析の結果、ロールバック数の符号は正であり、彼らの仮説に反する。一方バージョン数、ユーザー数、タグ修正回数のうち、前2者は正の、後者は負の符号を有し、かつ1%水準で統計的に有意であり、彼らの仮説と合致する。以上から、主題正確度に関しては、5指標のうちバージョン数、ユーザー数、タグ修正回数が信頼性を表す指標であると考えられる。

4. おわりに

本稿では、公共データとVGIを比較し、対象とするVGIでISO上のデータ品質が担保されていることを確認した上で、VGIのデータ品質保証に関して、Keßler and de Groot (2013)が提示した5つの信頼性指標の有用性を明らかにすることを目的とした。結果として、VGIとして取り上げたOSMの位置・主題正確度はともに高かったことをまず明らかにした。しかし重回帰分析の結果から、5指標を、VGIのデータ品質を保証するための信頼性を表す指標であるとはいいたいことも示した。これらの指標のうちバージョン数、ユーザー数、タグ修正回数の3指標が主題正確度の信頼性を表す指標となりうると考えられる。

この結果を踏まえて、信頼性を表す新たな指標を開発する余地があることと、信頼性以外のreputationといった指標を用いて、VGIのデータ品質を実証的に評価する必要があるといえる。さらに信頼性指標単独で、VGIのデータ品質を担保しがたいことが本研究で示唆されたことから、外部評価者にもとづくソーシャル・アプローチの有用性が検討されるべきであるとも考えられる。

謝辞

本研究を遂行するにあたり、国土地理協会、2019年度学術研究助成(申請者:山下潤)の一部を使用した。

参考文献

- 金杉洋・瀬戸寿一・関本義秀・柴崎亮介 (2019) オープンストリートマップ道路データとデジタル道路地図の比較—位置と完全性に着目して—。「GIS—理論と応用」, **27**, 43-48.
- Bishr, M. and Kuhn, W. (2007) Geospatial information bottom-up: a matter of trust and semantics. In Fabrikant, S.I. and Wachowicz, M. eds. *The European information society, Lecture Notes in Geoinformation and Cartography*, Springer, Berlin, Heidelberg, 365-387.
- De Longueville, B., Luraschi, G., Smits, P., Peedell, S. and De Groeve, T. (2010) Citizen as sensors for natural hazards: A VGI integration workflow. *Geomatica*, **64**, 41-60.
- Flanagin, A.J. and Metzger, M.J. (2008) The credibility of volunteered geographic information. *GeoJournal*, **72**, 137-148.
- Goodchild, M.F. and Li, L. (2012) Assuring the quality of volunteered geographic information. *Spatial Statistics*, **1**, 110-120.
- Haklay, M. (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, **37**, 682-703.
- Haklay, M., Basiouka, S., Antoniou, V. and Ather, A. (2010) How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, **47**, 315-322.
- Keßler, C. and De Groot, R.T.A. (2013) Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In Vandenbroucke, D., Bucher, B.

- and Cromptoets, J. eds. *Geographic Information Science at the Heart of Europe*, Springer, Berlin, Heidelberg, 21-37.
- Mezzetti, N. (2004) A Socially Inspired Reputation Model. In Katsikas, S. K., Gritzalis S. and López J. (eds.) *Public Key Infrastructure. EuroPKI 2004. Lecture Notes in Computer Science, 3093*, Springer, Berlin, Heidelberg, 191-204.
- Moreri, K. K., Fairbairn, D., James, P. (2018) Volunteered geographic information quality assessment using trust and reputation modelling in land administration systems in developing countries. *International Journal of Geographical Information Science*, **32**(5), 931-959.
- Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. and Haklay, M. (2017) A Review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, **31**, 139-167.
- Sztompka, P. (1999) *Trust: a sociological theory*. Cambridge University Press, Cambridge.
- Yamashita, J., Seto, K., Nishimura, Y. and Iwasaki, N. (2019) VGI contributors' awareness of geographic information quality and its effect on data quality: A case study from Japan, *International Journal of Cartography*, **5**, 214-224.