

OpenStreetMap の貢献者のコミュニティ活動による生産性の分析

Analysis of Productivity by Community Activities of Contributors in the OpenStreetMap

早川 知道^{*1*2} 伊藤 孝行^{*2}
Tomomichi Hayakawa Takayuki Ito

^{*1} オープンストリートマップ ファンデーション ジャパン
OpenStreetMap Foundation Japan

^{*2} 名古屋工業大学
Nagoya Institute of Technology

The OpenStreetMap is a geographic information project by a voluntary user's participation. Productivity of the OpenStreetMap was analyzed in community contributors. In an early stage, some contributors were very active but the number of them is small. In addition, at a certain point, contributors, whose productivity is small, begins to increase. Based on this certain point, we found that successful projects show an U-shaped curve in terms of the number of participants and productivity. Also, by regional events or contributors activities, it was found that in some cases k value varies.

1. はじめに

本稿では、ユーザー参加型によるボランタリーな地理情報作成プロジェクトである OpenStreetMap^{*1} において、貢献者の共同体であるコミュニティにおける生産性について分析を行う。特に、OpenStreetMap の地域コミュニティの各貢献者が作成した成果物数のランキングの構成が、貢献者の増加により変化する様子を調査分析する。

調査分析を行った結果、新たに次の事が分かった。

- 地域の活動の初期段階から発展段階にかけて、コミュニティの生産性の傾きを示す k 値が徐々に小さくなった。つまり、貢献者の少ない活動の初期には一部の貢献者が活発に活動を行っていたが、貢献者が増えるに従い多くの貢献者が活発に活動を行うようになった事を示している。
- ある時点を基点として、k 値が大きくなっていく事により、U 字型のカーブを描く事が分かった。つまり、ある時点より生産性の少ない貢献者が増えるようになった事を示している。
- 地域的な事象によるイベントや貢献者の活動状況によっては、一時的に k 値が逆方向に変動する場合もある事も分かった。

本稿の構成は、次の通りである。2 章では、本稿の目的と背景について述べる。3 章では、OpenStreetMap の概要とデータの仕様について説明する。4 章では、調査に用いたデータの取得方法と抽出方法について説明する。5 章では、本稿の調査結果を述べる。6 章では、調査結果を基に考察を行う。7 章では、本稿で得られた知見をまとめ、今後の研究課題を示す。

2. 目的と背景

Almeida らの研究 [1] では、Wikipedia^{*2} の発足から発展における貢献者と成果物の増加について、成果物 (記事) と貢献者

の数の累計変化を分析し、Wikipedia の発展傾向を明らかにしている。Almeida らは、Wikipedia の成果物の編集数の貢献者ランクにおいて、Zipf の法則 [2] の式 (1) を用いて、Wikipedia の貢献者らの生産性について分析を行っている。Zipf の法則とは、頻度が r 番目の要素は、頻度が 1 番目の単語の $1/r$ の確率で現れるという法則である。分析では、Zipf の法則の曲線の k 値が小さい編集数が多い貢献者の上位グループと、k 値が大きい貢献者の下位グループとの、2 つのグループに分かれる事を示している。2 つのグループに分かれる事により、k 値が小さい上位グループの方が生産的であると述べている。

$$p(r) \propto 1/r^k \quad (1)$$

$p(r)$: 貢献者 (r) の成果物数
r: 貢献者ランク (作成した成果物数の多い順)
k: 曲線の傾き

本稿では、ボランタリーな貢献者のコミュニティにおける、発展過程の生産性の変化に着目して調査を行う。具体的には、OpenStreetMap の各地域コミュニティ毎に、生産性の変化について Zipf の法則を用いて調査分析を行う。また、複数の地域コミュニティの変化を調査する事で、知見の一般化を目指す。さらに、日本の OpenStreetMap の早期の発展の為に、OpenStreetMap 先進地域の発展、及び普及状況を調査分析し、得られた知見および問題点を、OpenStreetMap コミュニティへフィードバックする。

貢献者らの活動は、活動の初期段階から、活性化し発展する過程において、多くの貢献者が参加し、貢献者のコミュニティ自体が変化して成長していく。貢献者のコミュニティ自体が変化するのであれば、成果物を作成する形態も変化する。地域のコミュニティ単体で、どのように変化していくのか調査し、更に、複数の地域コミュニティでの共通点等を調査する。

k 値の変化と貢献者のコミュニティの活動について、次にまとめる。k 値が大きい時は、一部の貢献者が多くの成果物を作成している状態、若しくは貢献者数が少ない場合である。成果物作成数の多い貢献者と、成果物作成数の少ない貢献者との、成果物作成数の差が大きき状態である。ボランタリーなコミュニティにおいては、一部の貢献者に負担が集中し持続可能な活動が困難な状態と考える。活動の初期に多く、貢献者を増やす活動なども必要である。

連絡先: 早川 知道, 名古屋工業大学 伊藤孝行研究室, 名古屋市昭和区御器所 名古屋工業大学 19 号館 205 室, 052-735-7968, hayakawa.tomomichi@nitech.ac.jp

*1 OpenStreetMap, <http://openstreetmap.org/>

*2 Wikipedia, <http://wikipedia.org/>

逆に, k 値が小さい時は, 多くの貢献者が比較的同じように成果物を作成している状態であり, 一部の貢献者が飛び抜けて多くの成果物を作成していない状態である. 成果物作成数の多い貢献者と, 成果物作成数の少ない貢献者との, 成果物作成数の差が小さい状態である. ボランティアなコミュニティにおいては, 一部の貢献者に負担が集中する事が無い為, 持続可能な活動が可能であり, 理想的な形態であると考えられる.

3. OpenStreetMap

3.1 OpenStreetMap について

OpenStreetMap とは, 世界中の様々な地理情報に基づく周知情報を集約したデータベースを作成する, ユーザー参加型によるボランティアなプロジェクトである. データは, オープンソースと同様の Open Data Commons Open Database License (ODbL) ^{*3} のライセンスにより, 誰でも自由に利用する事が出来る. OpenStreetMap のプロジェクトは, 2004 年に Steve Coast 氏 [3] により始められた. 著作権等の法的問題や技術的問題などから自由に地図が使えないケースが多く, 創造的または生産的な地図の利活用の促進の手段として始められた. OpenStreetMap は, 全世界で約 153 万人 (2014 年 3 月現在) ^{*4} がユーザーとして登録され, データ編集ユーザー (貢献者) を「マッパー」と称し, 成果物である世界各地の地理情報データベースの作成や更新を行う. 編集作業はクラウドソーシングで行われ, Wikipedia 同様に複数の貢献者による同時編集作業が可能である.

3.2 OpenStreetMap のデータの仕様

OpenStreetMap のデータは, オブジェクト (Object), 及びタグ (Tag) により構成され, 編集履歴 (Change Set) により全編集履歴が管理されている. OpenStreetMap のデータは, オブジェクトに対してタグによる情報を付加する仕様となっている. 本稿ではオブジェクトを成果物として調査分析した.

オブジェクトは, ノード (Node), ウェイ (Way), 及びリレーション (Relation) の 3 種類から成る. ノードとは点の事である. ノード単体で地点情報の登録に用いる場合と, ウェイの構成単位となる場合がある. また, 緯度経度の情報はノードについて与えられるので, 位置の情報を得る事が出来る. ウェイとは線の事である. 道路, 線路若しくは河川等の中心線に用いる. また, 始点と終点在同一の閉じたウェイをエリア (Area) として, 建物, 河川の流域等の表現に用いる. 位置情報をもつノードを繋げる事により, ウェイが構成されるので, 形状の情報を得る事が出来る. リレーションとは, ノードやウェイをグループ化したものである. 交通機関の路線や複数棟の建物など複数の要素をまとめて表現する場合に用いる. また, 穴の開いた複雑な形状のエリア (マルチポリゴン: Multi-polygon) を表現する為にも用いる. つまり, オブジェクトは位置や形状の情報を記録する為に用いられる. オブジェクト自体は緯度・経度情報しか持たず, より多くの情報を持たせる為にタグを付加する.

タグは Key-Value 形式であり, オブジェクトに対して柔軟なタグ付けにより位置情報以外の様々な情報を付加するために用いる. Key-Value 形式とは, キー (Key) と値 (Value) の組み合わせでデータを保存する方法で, 設定したキー (Key) を指定すると, 対応した値 (Value) を取り出すことができる.

4. データの取得と調査手法

4.1 データの取得と抽出

OpenStreetMap の全球データ ^{*5} ^{*6} を 2008 年から 2012 年の期間について取得し, 各地域のデータを時系列で抽出し利用した. データの抽出には, OpenStreetMap データを処理する為のコマンドライン Java アプリケーションである osmosis [5] を用いた. 取得した OpenStreetMap の全球データ, および抽出した各地域のデータの総容量は約 3TB である.

さらに, 時系列で抽出した地域毎のデータにより, 貢献者毎に成果物数 (オブジェクト数) を集計し, 貢献者ランクデータを出力した. 貢献者ランクデータを地域毎に分析し, Zipf の法則の曲線の傾きである k 値を集計した.

データの処理には, 8Core CPU, およびメモリ 16GB の計算機を用い, Ubuntu13.10 の OS により処理を行った. 上記, OpenStreetMap の全球データの取得, 各地域データの抽出, および貢献者ランクデータの集計の作業に約 15 週間で完了した.

4.2 調査方法

図 1 及び図 2 は, 貢献者ランクをグラフにした例である. X 軸は, 貢献者ランクであり, 各貢献者が作成した成果物の多い順に並べてある. Y 軸は, 貢献者毎の成果物数である. また, X 軸と Y 軸共に, コミュニティ間の値の違いを吸収し正規化する為に, それぞれの値を最大値で割って, グラフの X 軸と Y 軸の最大値を 1 になるように調整した. 図 1 では, グラフの形がロングテールになっている. 図 2 では, 同じデータを対数表示にしたグラフである. 次に, Zipf の法則に基づく式 (1) を用い, 傾きである k 値を, 地域毎に複数の時期で取得し調査を行った.

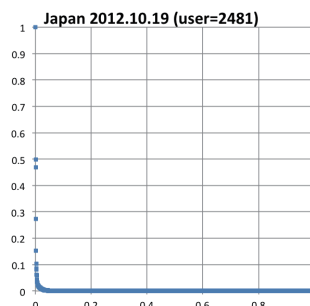


図 1: 貢献者ランクの例 (日本:2012 年 10 月 19 日)

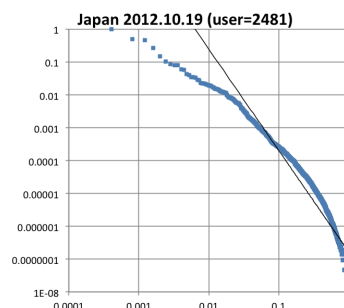


図 2: 貢献者ランクの例 (日本:2012 年 10 月 19 日) 対数表示

^{*3} Open Data Commons Open Database License (ODbL), <http://opendatacommons.org/licenses/odbl/>

^{*4} OpenStreetMap stats report, http://www.openstreetmap.org/stats/data_stats.html

^{*5} '全球' とは, 一般に世界若しくは地球の事をさす. '全球データ' とは, 世界中のデータである事を意味する.

^{*6} Planet OpenStreetMap, <http://planet.openstreetmap.org/>

5. 調査結果

5.1 日本の活動

図3は、日本の貢献者ランクの曲線の傾きであるk値を貢献者数の増加に合わせて表したグラフである。表1は、図3のグラフのデータである。

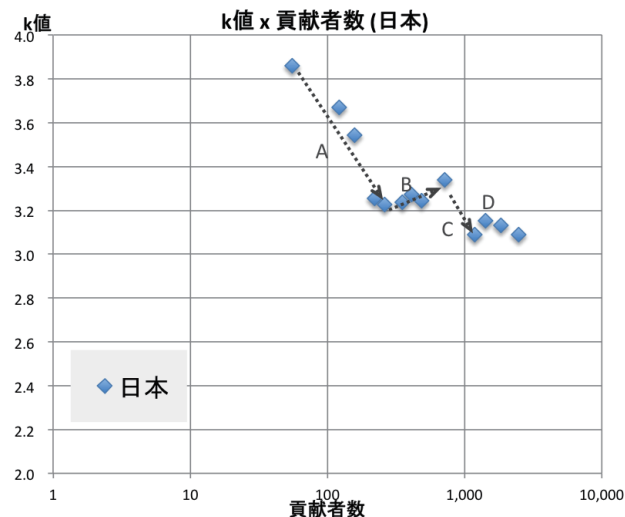


図3: k値と貢献者数(日本)

表1: k値と貢献者数(日本)

日付	貢献者数	k 値
2008/4/2	55	(a) 3.861
2008/10/1	122	3.670
2009/1/7	157	3.543
2009/4/21	220	3.254
2009/7/1	259	(b) 3.229
2009/10/7	352	3.237
2010/1/6	412	3.271
2010/4/1	484	3.246
2010/10/6	718	(c) 3.339
2011/4/6	1184	(d) 3.090
2011/10/5	1,415	(e) 3.154
2012/4/1	1,842	3.130
2012/10/19	2,481	3.088

図3のA(表1のaからb)は、初期の日本の活動時期であり、貢献者が増えるに従って、k値が小さくなる事が分かる。

図3のB(表1のbからc)は、一部の貢献者により大量のデータ(国土数値情報データ[7][8]の入力作業が行われた時期)である。一部の貢献者の大量の成果物の作成により、k値が大きくなっていく事が分かる。

図3のC(表1のcからd)は、2011年の東日本大震災直後に被災状況をデータ化する為のクライシスマッピング[10][11][12][13]が行われ、新たに多くの貢献者が参加し、多くの成果物を作成した時期である。急激にk値が小さくなっており、活発に成果物を作成する貢献者が急激に増加した事が分かる。

図3のD(表1のe)は、クライシスマッピングの作業を補完する為に、Yahoo道路データ[9]の入力が行われた時期であ

る。一部の貢献者により、Yahoo道路データの入力が行われており、一時的にk値が大きくなっている事が分かる。

以上により、日本においては、一部の貢献者が大量のデータ入力を行ったケースや、クライシスマッピングにより集中的に大勢の貢献者が成果物を作成したケースにより、一時的にk値が変動した事が分かる。日本のk値の変動は、地域的な事象によるイベントや貢献者のイレギュラーな活動による、小規模な変動である。従って、日本のk値は、初期の活動時期より下降傾向である事が分かる。

5.2 OSM 先進地域の活動

次に、OSM 先進地域である英国、ドイツ、オランダ、フランス、及びスペインの調査を行った。図4は、英国、ドイツ、オランダ、フランス、及びスペインの貢献者ランクの曲線の傾きであるk値を貢献者数の増加に合わせて表したグラフである。

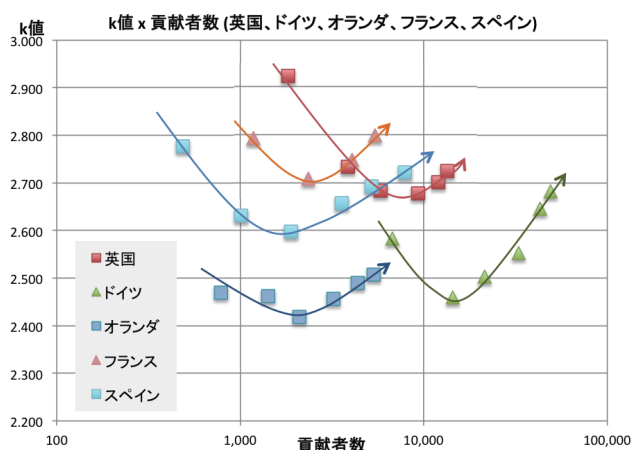


図4: k値と貢献者数(英国、ドイツ、オランダ、フランス、及びスペイン)

全ての地域において、貢献者数が増加するに従いk値が小さくなっていったが、ある時期を境にk値が大きくなり、U字型となっている事が分かる。また、本稿の調査では、U字の底にあたる基点では、k値がおおよそ2.4~2.7である事が分かる。各地域において、ほぼ同様に見られるが、時期は異なる為、特定の事象が原因では無いと考えられる。貢献者が増えてコミュニティが活性化していたが、新たに生産性の低い貢献者参加するようになったと言える。

本稿の調査は2008年以降のデータである為、各地域の左辺側の活動の初期段階のデータは取得する事が出来なかった。しかし、どの地域も貢献者が少ない時期は下降傾向にあり、それ以前も同様の下降傾向にあったと想像する事は出来る。各地域の活動の初期段階のデータの調査は、今後の課題とする。

6. 考察

各地域の初期の活動時期から貢献者数が増加する事により、k値が小さくなる事を確認した。日本では、大量のデータ(国土数値情報データ、及びYahoo道路データ)入力作業や東日本大震災直後のクライシスマッピングが行われた影響により、一時的にk値の傾向は変動していたが、全体の傾向としてk値が小さくなる傾向である事が確認できた。OSM 先進地域においては、初期のデータが不足した為、確認出来なかったが、取得出来たデータにおいてはk値が下降傾向にあり、基本的には下降

傾向である。不足していたデータの調査については、今後の課題とする。

活動が進み、貢献者が増加すると、逆に k 値が大きくなる事を確認した。本稿で調査を行った OSM 先進地域では、全ての地域において、 k 値が小さくなる傾向から大きくなる傾向に変わっており、U 字型になっていた。日本においても、 k 値が小さくなる傾向から大きくなる傾向が見られたが、一時的な変動であり、特定の事象である事が原因と分かっている。また、OSM 先進地域では、U 字の底は k 値が 2.4 から 2.7 の間を基点としていたのに対して、日本は k 値が 3.0 以上で傾向が変わっていた。よって、日本においては、 k 値が小さくなる傾向から大きくなる傾向になるような、U 字型のカーブは確認出来なかった。

地域の活動状況により、 k 値は次のように変化する。発展の初期段階においては、地域の活動のはじまりであり、貢献者数は少ないが一部の活発な貢献者が中心となり活動が行われる。クリティカルマス [6] で言われる、イノベーター層の活動とも言える。結果として、地域全体の k 値は大きい状態から始まる事になる。初期の貢献者の活動の成果により新たな貢献者が参加する事になり、活動の発展が進む。貢献者が増える事により、活発な貢献者が増える。結果として、地域全体の k 値が徐々に小さくなっていく。更に、活動の発展が進み貢献者が増えてくると、今度は、あまり活発でなく生産性の低い貢献者が増加する。結果として、ある時点を経験して、地域全体の k 値は大きくなっていく。従って、 k 値は貢献者の増加に対して U 字型のカーブを描く事になる。また、一部の貢献者が大量のデータ入力を行った場合や、クライシスマッピング等のように集中的に大勢の貢献者が成果物を作成するような場合は、一時的に k 値が逆方向に変動する場合がある。

OSM 先進地域では、今後も k 値が大きくなり続ける事は、一部の特定の貢献者の活動だけが活発であり続ける事であり、貢献者が増えたとしても、多くは生産的でない貢献者である。持続可能なコミュニティ活動としては理想的な状態では無い。

k 値が小さい状態が、貢献者毎の負担が小さく持続可能なコミュニティとして理想的であるので、OSM 先進地域も今後は、再び k 値が小さくなる、若しくは一定の k 値で安定する可能性がある、と考えられる。

つまり、 k 値の傾向が変化するのには、日本のように特定のイベントによるケースと、OSM 先進地域のように貢献者の活動形態による自然発生的なケースがあると考えられる。 k 値の傾向の変化を繰り返し、持続可能なコミュニティ活動として理想的な状態に近づいていくためには、今後も観察が必要である。

7. まとめ

本稿では、ユーザー参加型によるボランティアなプロジェクトである OpenStreetMap において、貢献者の共同体であるコミュニティにおける生産性について分析を行った。OpenStreetMap の地域コミュニティの各貢献者が作成した成果物数のランキングの構成が、貢献者の増加により変化する様子を調査分析した。調査分析の結果、次の事が分かった。

- 地域の活動の初期段階から発展段階にかけて、コミュニティの生産性の傾きを示す k 値は徐々に小さくなっていく。
- ある時点を経験して、 k 値が大きくなっていく事により、U 字型のカーブを描く事が分かった。
- 地域的な事象によるイベントや貢献者の活動状況によっては、一時的に k 値が逆方向に変動する場合もある事も分かった。

今後の課題を述べる。各地域の初期のデータ及び最新のデータを取得する事、及び調査地域を増やす事により、更に詳細に分析を行う必要がある。貢献者ランクにおいて、複数の k 値を用いてコミュニティの生産性を詳細に計ると共に、複数の k 値を用いた変化について分析を行う。更に、調査を深め知見を一般化する。

参考文献

- [1] Almeida, R. B., Mozafari, B. and Cho, J. : On the Evolution of Wikipedia, International Conference on Weblogs and Social Media (ICWSM'07), (2007)
- [2] G. K. Zipf. : Human Behavior and the Principle of Least Effort, Addison-Wesley (Reading MA), (1949)
- [3] Coast, S.: How OpenStreetMap is changing the world, In proc. of 10th International Symposium on Web & Wireless GIS(W2GIS2011), pp4, (2011)
- [4] Planet OpenStreetMap, <http://planet.openstreetmap.org/>, 2014.3.5
- [5] Osmosis, <http://wiki.openstreetmap.org/wiki/Osmosis>, 2014.3.5
- [6] Rodger, E.: "Diffusion of innovations, Fifth Edition" , Simon and Schuster, (1995)
- [7] 国土政策局 国土数値情報 ダウンロードサービス, http://nlftp.mlit.go.jp/ksj/jpgis/jpgis_datalist.html, 2014.3.5
- [8] Japan KSJ2 Import, <http://wiki.openstreetmap.org/wiki/Import/Catalogue/Japan.KSJ2.Import>, 2014.3.5
- [9] JA:YahooJapanALPS Data/Highway, http://wiki.openstreetmap.org/wiki/JA:YahooJapanALPS_Data/Highway, 2014.3.5
- [10] 2011 Sendai earthquake and tsunami, http://wiki.openstreetmap.org/wiki/2011_Sendai_earthquake_and_tsunami, 2014.3.5
- [11] 瀬戸寿一: 災害対応におけるボランティアな地理空間情報の時空間的推移:東日本大震災クライシス・マッピング・プロジェクトを事例に, 地理情報システム学会講演論文集, (2011)
- [12] 伊美裕麻, 早川知道, 伊藤孝行: 震災時における OpenStreetMap の利用と推移に関する考察, ネットワークが創発する知能研究会 (JWEIN'12), August 29-31,2012, pp.8, (2012)
- [13] Yuma Imi, Tomomichi Hayakawa and Takayuki Ito: Analyzing Effect of OpenStreetMap for Crisis, The 1st International Workshop on Smart Enterprise and Mobile Platforms (SEMP2012), The 14th IEEE International Conference on Commerce and Enterprise Computing (CEC), Hangzhou, China, September 9-11,2012, pp.5, (2012)