

貢献者と成果物の増加率変化から見る OpenStreetMapの発展に関する考察

伊美 裕麻*¹ 早川 知道*¹ 伊藤 孝行*¹
Yuma Imi Tomomichi Hayakawa Takayuki Ito

*¹名古屋工業大学

Nagoya Institute of Technology

The goal of this paper is to clarify changes of the number of contributors and user-generated content of OpenStreetMap, and discuss its development. In a new generation of IT-enabled collaborative tools, such as Wikipedia, an excessive number of contributors negatively influence the value of user-generated content. Data of OpenStreetMap is used in many social activities, so it's necessary for us to understand more about influence of contributors to results in OpenStreetMap. Our main findings include that (1) the number of contributors had been growing linearly, and (2) the number of information had shown growth curve. The increasing rate decelerate as it approaches the limit

1. はじめに

コラボレーションツールでは、貢献者数の変化が成果物に対して大きく影響する。例えば、コラボレーションツールの1つである Wikipedia では、貢献者数の指数関数的な大きな増加に伴い、成果物数が大きく増加してきた [1]。

しかし、現在のコラボレーションツールでは、単純に貢献者が増加すればよいわけではない。Wikipedia では、貢献者は増加し続けているものの、成果物の編集活動はゆるやかになり、安定しはじめている [2]。さらに、貢献者の過大な増加は成果物の品質に対して悪影響を及ぼす可能性があることが指摘されている ([3], [4])。今後のコラボレーションツールの発展には、貢献者の増加の影響を理解する必要がある。

新たなコラボレーションツールの1つに、OpenStreetMap [5] がある。OpenStreetMap は 2004 年にイギリスの Steve Coast*¹氏によりはじめられた自由な地図の作成を目的としたプロジェクトであり、地図版の Wikipedia と呼ばれる [6]。現在 OpenStreetMap のデータは商用、防災、教育、福祉、産業、および観光などのさまざまな分野に活用されており、再利用可能なデータに注目が集まっている。

しかし、実際に成果物が充実しているのは主要都市を中心とした一部であり、まだ発展途中である部分が多く存在する ([7], [8])。今後利用の期待される部分をより効率的に発展させることで、数少ない自由な地理情報をより円滑に充実させ、多くのサービスおよび取り組みを活性化させることが期待される。

そこで本論文では、先行しているコラボレーションツールである Wikipedia での変化を参考に、増加率の変化から OpenStreetMap の貢献者数と成果物数の成長過程を明らかにすることを目的とする。本論文の結果を得ることで、OpenStreetMap 全体、および一部の今後の変化の理解を促す。

本論文の構成は以下の通りである。第2章では、Wikipedia や OpenStreetMap の先行研究で見られる傾向から、貢献者数と成果物数の変化について議論する。第3章では、データ処理による分析手法を示す。第4章では、分析の結果を示し、議論を行い、最後にまとめを示す。

2. 貢献者数と成果物数の変化の検討および仮説

2.1 貢献者数

OpenStreetMap の貢献者数は指数関数的に増加すると考える。Web 上のコラボレーションツールにおいて、アカウント登録を行なった利用者を登録者と呼ぶ。貢献者は、登録者のうち実際に情報の作成、及び編集を行なった者と定義する。

Wikipedia では、貢献者数は指数関数的に増加していくことが分かっている [1]。Wikipedia は現在も利用規模を大きく発展させている。また、OpenStreetMap では、登録者数に関しては Wikipedia と同様に指数関数的に増加している [8]。また、OpenStreetMap のデータを集計した OpenStreetMap Stats [9] でも登録者数に同様の傾向が見られている。OpenStreetMap の登録者数は、プロジェクトの発足以来大きく増加している。以上から、OpenStreetMap の貢献者数の変化について仮説を立てる。

【仮説1】 貢献者数は指数関数的に増加する。

2.2 成果物数

発展の進むはじめの段階では、成果物数は指数関数的に増加する。しかし、貢献者の増加により成果物の増加は停滞し、指数関数的ではなくなると考える。OpenStreetMap の編集では、まずはじめに衛星写真などを用いて編集したい地理情報の形状を編集する。地理情報の形状や緯度経度などの基本的な情報を OpenStreetMap ではオブジェクト (Object) と呼ぶ。OpenStreetMap の成果物とは、地図を作成するのに必要なオブジェクトである。本論文では、成果物数としてオブジェクト数の変化について議論する。

Wikipedia では、貢献者数と同様に成果物数も指数関数的に増加していくことが分かっている [1]。また、OpenStreetMap においても全体の成果物数は Wikipedia と同様に指数関数的に増加している [9]。しかし、より近年では Wikipedia において貢献者が増加し続ける中で、編集数の増加は止まり、成果物への編集数は一定になると指摘されている [2]。つまり、編集活動の拡大には限界があるといえる。

OpenStreetMap の成果物について考える。OpenStreetMap の成果物は、地上に存在する建物や道などをデータとして作成する。つまり、成果物数には編集範囲に存在する建築物という上限値が存在する。

連絡先: 伊美裕麻, 名古屋工業大学 産業戦略専攻, 愛知県名古屋市中区御器所町, TEL:052-735-7968, FAX: 052-735-7407, MAIL: imi.yuma@itolab.nitech.ac.jp

*¹ OpenStreetMap の創設者。2011 年から Microsoft に所属。

以上のことから OpenStreetMap についても、発展の進む段階では Wikipedia と同様に成果物数が指数関数的に増加する。しかし、貢献者数の大きな増加に伴い、編集活動が安定し、成果物数の増加が停滞することが考えられる。OpenStreetMap の成果物数の変化について仮説を立てる。

【仮説 2】 成果物数は指数関数的に増加する。しかし、貢献者の増加に伴い成果物の増加は停滞する。

本論文では上の 2 つの仮説の検証を行い、OpenStreetMap における貢献者数と成果物数の成長過程を明らかにする。

3. 大規模データ処理を用いた分析手法

3.1 分析対象の検討とデータ処理

OpenStreetMap の全体は、現在も規模を大きく発展させている ([8], [9])。本論文では、さらに発展の進んだ段階における変化を分析するため、OpenStreetMap の利用が特に活発であるドイツの国全体および地域を分析対象として選択した。

ドイツは OpenStreetMap の活動が活発であるヨーロッパの国々の中でも、貢献者数が顕著に多く、活動も活発である [8]。さらに、ドイツの中でも主要都市では、発展が進み成果物数が充実していることがわかっている [7]。本論文では、主要都市のうち Berlin, Hamburg, Koln, および Munchen の 4 都市を対象とする。

本研究では、世界データを OpenStreetMap の履歴データサーバー (planet.osm[10]) から取得した。世界データは 78ヶ月分約 15TB を取得した。世界データを利用し、細分化の処理を行うことで、国および地域規模のデータを作成した。データの細分化には osmosis を利用した。osmosis とは OSM データを処理するコマンドライン Java アプリケーションである。本研究のデータの細分化とは、一部の範囲のデータをサブセットとして作成することである。全体のデータを作成するのに約 3.5 週間を要した。以上のデータを用いて貢献者数、および成果物数の集計を行なった。

3.2 分析方法

本論文の目的は OpenStreetMap の貢献者数および成果物数の成長過程を明らかにすることである。変化傾向の分析のため、貢献者数と成果物数の時間経過の変化について回帰分析を行う。

貢献者数には線形モデルと指数モデルへの回帰分析を行なった。線形モデルには式 (1)、また指数モデルには (2) を用いた。式 (1) および式 (2) は時間 t の関数である。

$$y = at + b \quad (1)$$

$$y = ae^{bt} \quad (2)$$

成果物数には線形モデル、指数モデル、および成長曲線モデルであるロジスティック曲線とゴンベルツ曲線への回帰分析を行なった。

成長曲線モデルについて説明する。成長曲線とは、任意の個体累計数の変化を表す S 字型の曲線で、時間が経つにつれて増加が止まり、個体数が一定値に近づく様な曲線である。つまり、上限値に近づくにつれて増加が停滞していくモデルである。成長曲線は人口増加など上限が存在する個体増加のモデルとして作成された。

本論文では、成長曲線モデルには代表的な 2 つであるロジスティック曲線とゴンベルツ曲線を用いた。成長曲線の例を図 1

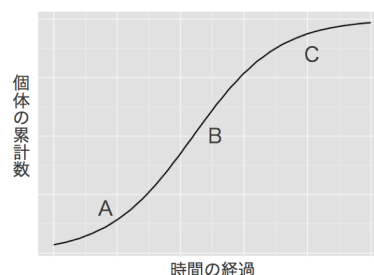


図 1: 成長曲線の例

に示す。図 1 の A から B の段階では、増加率が上昇する。上昇は指数関数に近い。B から C の段階では上限値に向かい増加率が低下し、増加が停滞する。

ロジスティック曲線は式 (3) で表される。式 (3) 中の数値 K, b , および c により曲線の式を決定する時間 t の関数である。ゴンベルツ曲線は式 (4) で表される。各係数 K, b, c , および t はロジスティック曲線と同様の役割を果たす数値である。

$$y = \frac{K}{1 - be^{-ct}} \quad (3)$$

$$y = Kb^{e^{-ct}} \quad (4)$$

以上のモデルへの回帰分析を行うことで成果物数の変化を検証する。またモデル間の当てはまりの比較には AIC

(赤池情報量基準)[11] を用いた。AIC は値が小さいほど当てはまりがよいことを示す。AIC により最適なモデルを選択する。

4. 分析結果

4.1 結果：ドイツの 4 地域

ドイツの 4 地域 (Berlin, Hamburg, Koln, および Munchen) について分析の結果を示す。

ドイツの 4 地域の貢献者数

【仮説 1】 貢献者数は 指数関数的 に増加する。
【結果 1】 貢献者数は 線形的 に増加する。

貢献者数を分析した結果、ドイツの 4 地域 (Berlin, Hamburg, Koln, および Munchen) 全てに線形的な増加傾向が見られた。貢献者数の線形的な増加傾向は仮説 1 とは異なる結果である。

表 1 に回帰分析の結果による AIC を示す。全ての地域において指数モデルより線形モデルに当てはまりがよい。

表 1: AIC によるモデル比較 (貢献者数)

Area	線形	指数
Berlin	372	460
Hamburg	311	377
Koln	355	406
Munchen	383	442

貢献者数の変化を図 2 に示す。図 2 は、時間の経過に対する貢献者数の増加を示している。ドイツの 4 地域をそれぞれ●が Berlin, ▲が Hamburg, ■が Koln, および+が Munchen の貢献者数を示している。

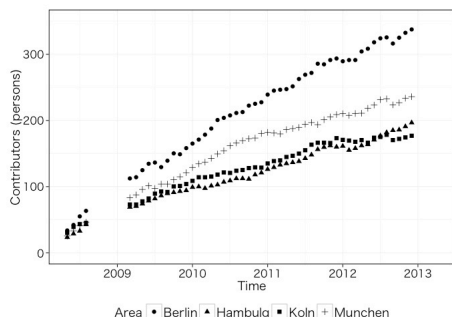


図 2: ドイツ 4 地域の貢献者数の累計

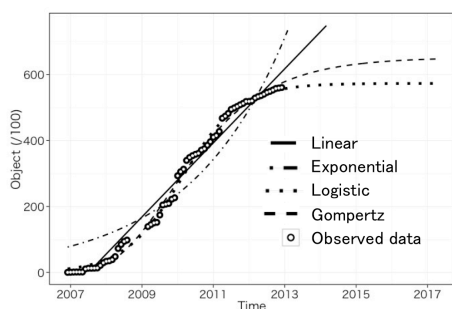


図 3: Berlin のモデル曲線

図 2 を見てもドイツの 4 地域 (Berlin, Hamburg, Koln, および Munchen) の貢献者数は、4 地域全てにおいて時間経過とともに線形に増加する。

ドイツの 4 地域の成果物数

【仮説 2】成果物数は指数関数的に増加する。しかし、貢献者の増加に伴い成果物の増加は停滞する。

【結果 2】成果物は成長曲線的に増加傾向が変化する。また、成果物の増加は停滞する。

成果物数を分析した結果、仮説 2 を検証することができた。ドイツの 4 地域 (Berlin, Hamburg, Koln, および Munchen) の全ての成果物数は成長曲線的な増加傾向であった。

表 2 に回帰分析の結果による各モデルの AIC を示す。各地域において最も当てはまりの良いモデルを太字で示す。全ての地域で指数モデルより成長曲線モデルへの当てはまりがよい。Berlin, Hamburg, および Munchen の成果物数はロジスティック曲線に、Koln の成果物数はゴンベルツ曲線に最も当てはまる。

表 2: AIC によるモデル比較 (成果物数)

Area	線形	指数	ロジスティック	ゴンベルツ
Berlin	658	795	546	557
Hamburg	785	839	616	619
Koln	638	687	524	493
Munchen	837	759	692	716

例として当てはめたモデルの近似線を Berlin を図 3, および Koln を図 4 に示す。以上の図は、時間の経過に対する成果物数の増加を示している。

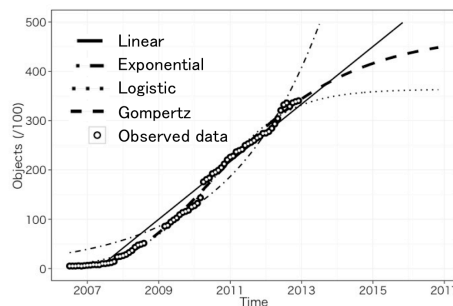


図 4: Koln のモデル曲線

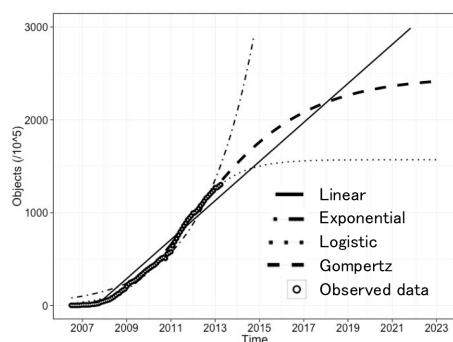


図 5: ドイツの成果物数のモデル曲線

Berlin にはロジスティック曲線が当てはまる。また実測値と近似線を見ても、成果物数の増加は明らかに停滞しはじめている。また、Koln にはゴンベルツ曲線が当てはまる。Koln は、実測値が成長曲線の途中段階であるが、増加率は低下している。

4.2 結果：国単位のドイツ

次に、国単位のデータであるドイツについて分析の結果を示す。

ドイツの貢献者数

【仮説 1】貢献者数は指数関数的に増加する。

【結果 1】貢献者数は線形的に増加する。

貢献者数を分析した結果、線形的な増加傾向が見られた。貢献者数の線形的な増加傾向は仮説 1 とは異なる結果である。しかし、ドイツの 4 地域の傾向とは同様である。表 3 に回帰分析の結果による AIC を示す。貢献者は指数モデルより線形モデルに当てはまりがよい。

表 3: AIC によるモデル比較 (貢献者数)

Area	線形	指数
Germany	372	460

ドイツの成果物数

【仮説 2】成果物数は指数関数的に増加する。しかし、貢献者の増加に伴い成果物の増加は停滞する。

【結果 2】成果物は成長曲線的に増加傾向が変化する。また、成長曲線の中期段階である。

成果物数を分析した結果、仮説2を検証することができた。ドイツの成果物数は成長曲線的な増加傾向であった。つまりドイツの4地域の傾向と同様である。

表4に回帰分析の結果によるAICを示す。最も当てはまりの良いモデルを太字で示す。ドイツの成果物数は指数モデルより成長曲線モデルへの当てはまりがよい。AICにより、ゴンペルツ曲線に最も当てはまる。

表4: AICによるモデル比較(成果物数)

Area	線形	指数	ロジスティック	ゴンペルツ
Germany	996	964	792	762

当てはめたモデルの近似線を図5に示す。実測値と近似線を見てもゴンペルツ曲線がよく当てはまっている。またドイツの成果物は図1のAからBに入った成長曲線の中期段階である。

4.3 議論

以上の結果から、考察を示す。

【仮説1】 貢献者数は指数関数的に増加する。

貢献者数の分析の結果、仮説1とは異なり、貢献者数の線形的な増加傾向が明らかになった。ドイツの4地域およびドイツの国全体の貢献者数はどちらも線形的な増加傾向であった。

OpenStreetMapの貢献者数の変化について検討する際、OpenStreetMapの登録者の指数関数的な増加を参考にした。しかし、分析の結果、登録者と貢献者の増加に差が大きくあり、利用者が編集に抵抗を持っている可能性を指摘する。Wikipediaの成果物はテキストであり、Web利用者の多くが利用したことがあるため、編集にも慣れている。それに対しOpenStreetMapの成果物は地図であり、多くの利用者が編集の経験がない。以上からOpenStreetMapの利用者には編集への抵抗があると考えられ、今後地図編集への抵抗を減らしていく必要があると考える。

【仮説2】 成果物数は指数関数的に増加する。しかし、貢献者の増加に伴い成果物の増加は停滞する。

成果物数の分析の結果、仮説2を検証することができた。ドイツの4地域およびドイツの国全体の成果物について成長曲線的な変化を明らかにした。

OpenStreetMapにおいても発展が進み成果物が充実することで、Wikipediaと同様に成果物数の増加が停滞する。今後、発展部分が広がることで、国やOpenStreetMap全体などのより広い範囲でも、成果物の増加が停滞することが考えられる。しかし、貢献者数が一方向に増加していることから、Wikipediaの先行研究で報告されている過大な貢献者の増加による悪影響([3],[4])を考えると、今後成果物の停滞する段階での貢献者の影響を明らかにする必要がある。

5. まとめと今後の課題

本論文では、OpenStreetMapにおける貢献者数と成果物数について回帰分析を行い、成長過程を明らかにした。地域および国のデータに着目し、発展した段階への変化を分析した。今後さまざまな部分の発展度合を示し、貢献者の活動が必要な部分を明示することにより、円滑で効率的な成果物の充実が期待される。OpenStreetMapの発展を促し、数少ない自由な地理情報が充実することで、現代のWebを利用した地理情報サービスやアプリケーションの向上、および新規サービスの可能性

の拡大が見込める。また、防災や福祉などの多くの分野で行われている取り組みへの貢献が期待される。

謝辞

本研究の一部は、内閣府の先端研究助成基金助成金(最先端・次世代研究開発プログラム)により助成を受けている。

参考文献

- [1] Rodrigo B. Almeida, Barzan Mozafari, Junghoo Cho: On the Evolution of Wikipedia, International Conference on Weblogs and Social Media (ICWSM'07) (2007)
- [2] Bongwon Suh, Gregorio Convertino, Ed H. Chi, Peter Pirolli: The Singularity is Not Near: Slowing Growth of Wikipedia, In Proceedings of WikiSym2009 (2009)
- [3] Gerald C. Kane and Sam Ransbotham: Collaborative development in Wikipedia, Collective Intelligence 2012 (2012)
- [4] Gerald C. Kane and Sam Ransbotham: Membership Turnover and Collaboration Success in Online Communities, Explaining Rises and Falls from Grace in Wikipedia MIS Quarterly 35(3), 613-627 (2011)
- [5] OpenStreetMap, <http://openstreetmap.org/>
- [6] Daniel Z. Sui: The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS, Computers, Environment and Urban Systems (2008)
- [7] Dennis Zielstra, Alexander Zipf: A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany, 13th AGILE International Conference on Geographic Information Science 2010 (2010)
- [8] Pascal Neis and Alexander Zipf: Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap, ISPRS Int. J. Geo-Inf. 146-165 (2012)
- [9] OpenStreetMap Stats, <http://wiki.openstreetmap.org/wiki/Stats>
- [10] OpenStreetMap planet.osm, <http://planet.openstreetmap.org>
- [11] Hirotugu Akaike: Information theory and an extension of the maximum likelihood principle, Proceeding of the Second International Symposium on Information Theory, Budapest, Akademiai Kiado, pp. 267-281. (1973)
- [12] Mordechai Haklay: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets, Environment and Planning B: Planning and Design 2010 (2010)