

MA4 生物多様性に対する関心のトレンドの Twitter マイニング

Twitter Mining for Detecting Interest Trends on Biodiversity: Messages from Seven Language Communities

指導教員 町村尚准教授・地球循環共生工学領域

2023/01/30 28H21008 石田愁 (Shu ISHIDA)

Abstract: IPBES has proposed a framework for sustainable future. Transformative change with respect to nature will be driven by recognizing what values people have; making inclusive decisions based on those values; restructuring policies, rights, and regulations in accordance with them; and transforming the social norms and goals that can drive change. Social media is a new source of information and a modern tool for monitoring public opinion on the human-nature interactions. This study identifies commonalities among seven language communities (those of the six official languages of the United Nations, plus Japanese) together with the showing the uniqueness of the Japanese community by comparing Tweets including the term biodiversity and to determining differences in interest and concern about biodiversity from the past to the present. Interest was then qualitatively and quantitatively identified using natural language processing technology. The diversity indexes have on the rise in all language communities but engagement behaved different by years and language communities. Network analysis showed the different understandings of biodiversity plays as the foundation of people's lives.

1. 背景・目的

社会変革をより持続可能で未来的にすることは、様々な自然の価値を中心とした介入点の組み合わせを実行できるかどうかにかかっており、このような中で人々の社会規範や目標を変えることが最も大きな効果をもたらすと言われている¹⁾。しかし、生物多様性に対する人々の関心や行動への意志を示す世界で共通して長期的に一貫した情報が不足している²⁾。そんな中、新しい人間と自然の相互作用に対するパブリックオピニオンの情報源やモニタリング手段として **Social media** が注目されている³⁾。よって本研究では、国連公用語と日本語を対象として、生物多様性に関連した **Tweet** を比較することで、言語コミュニティ間で共通している概念と日本語コミュニティの独自性を解析し、Twitter での生物多様性に対する人々の関心のトレンドを可視化することを目的とする。

2. 研究手法

2. 1 ツイートデータの取得

Tweet を取得するために、Twitter が公開する外部から Twitter サーバ内の Tweet を解析するためのインターフェースである Application Program Interface (API) として、Python 上で `tweepy` (ver 3.8.0)⁴⁾ を用いて Twitter API v2 を利用した。Tweet の取得期間は 2006 年 03 月 21 日から 2021 年 12 月 31 日までとし、Full archive 検索を利用した。検索ワードは '生物多様性' とし、日本語 ('生物多様性')、英語 ('biodiversity')、フランス語 ('biodiversité')、スペイン語 ('biodiversidad')、中国語 ('生物多样性')、ロシア語 ('биоразнообразие')、アラビア語 ('التنوع البيولوجي') の 7 言語コミュニティで Tweet を収集した。

2. 2 定量分析

エンゲージメントの評価は、年 i に取得したツイート数を $Tweet_i$ 、取得したツイートがリツイートされた合計を RT_i 、いいねされた合計を $like_i$ 、リプライされた合計を $reply_i$ とし、各年のエンゲージメント $RT_engagement_i$, $like_engagement_i$, $reply_engagement_i$ を式 (1), (2), (3) で算出した。また、年 i に利用される N 個のハッシュタグの確率分布 $p_1, p_2, p_3, \dots, p_n$ から算出されるシャノンエントロピーを $diversity\ index_i$ と呼び以下の式 (4) で算出した。

$$RT_engagement_i = \frac{RT_i}{Tweet_i} \quad (1)$$

$$like_engagement_i = \frac{like_i}{Tweet_i} \quad (2)$$

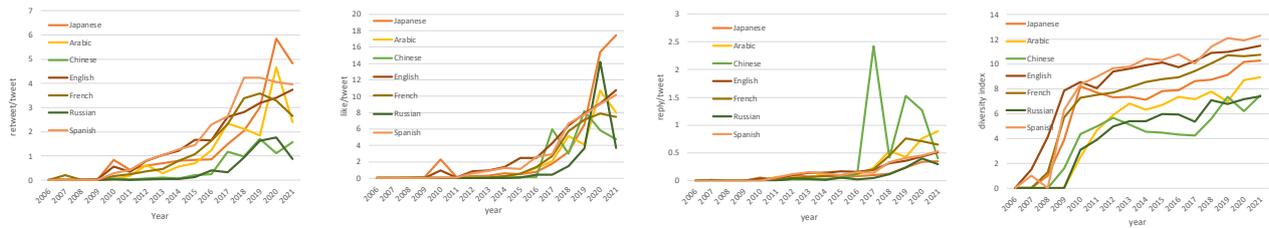
$$reply_engagement_i = \frac{reply_i}{Tweet_i} \quad (3)$$

$$diversity\ index_i = -\sum_{n=1}^N p_n \log_2 p_n [bit] \quad (4)$$

2. 3 定性分析

ハッシュタグの共起関係を調べるため Python 上で NetworkX (ver. 2.4)⁵⁾ を用いてネットワークグラフを作成し、媒介中心性を算出した。ネットワークグラフでは、ノードの大きさはハッシュタグの使用回数、エッジの太さはハッシュタグ間の共起回数に比例させた。また、あるノード i がそれ以外の任意の 2 つのノード s, t ($1 \leq s, t \leq n \cap s, t \neq i$) を結ぶ ${}_{n-1}C_2$ 個の最短経路 σ_{st} のうち、ノード i を通過する最短経路 $\sigma_{st}(i)$ の割合の和を媒介中心性 b_i として式 (5) で算出した。

$$b_i = \sum_{s=1}^n \sum_{t=1}^n \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (5)$$



a) RT_engagement b) like_engagement c) reply_engagement d) diversity index

Fig. 1 Engagement と diversity index の推移

Table 1 媒介中心性の高いハッシュタグ

	Japanese	Translated	between centers	Arabic	Translated	between centers	Chinese	Translated	between centers	English	between centers
1	biodiversity	biodiversity	0.37	قطر Diameter		0.14	COP15	COP15	0.37	Biodiversity	0.15
2	cop10	cop10	0.25	الكويت Kuwait		0.07	biodiversity	biodiversity	0.14	Wildlife	0.06
3	nagoya	Nagoya	0.13	أبوظبي Abu Dhabi		0.07	生物多样性 Biological diversity	Biological diversity	0.13	Nature	0.05
4	環境 surroundings	surroundings	0.13	الإمارات UAE		0.06	COP15inKunming	COP15inKunming	0.09	nature	0.04
5	kankyo	Quiet life	0.12	مصر Egypt		0.06	Conservation	Conservation	0.07	Environment	0.04
6	bdjp	BDJP	0.11	تغير المناخ Climate change		0.05	SDG15	SDG15	0.05	Conservation	0.02
7	news	news	0.09	اليوم العالمي للبيئة Today - the world_ of the environment		0.05	Connect2Earth	Connect2Earth	0.02	climatechange	0.02
8	COP10	COP10	0.08	السيسي Sisi		0.05	EndangeredSpecies	Endangered Species	0.02	conservation	0.01
9	生物 biology	biology	0.07	فرنسا France		0.04	EPIL	Hype	0.02	environment	0.01
10	nhk	nhk	0.07	شرم الشيخ Sharm EL-Shaikh		0.03	BRI	BRI	0.02	wildlife	0.01
	French	Translated	between centers	Russian	Translated	between centers	Spanish	Translated	between centers		
1	Nature	Nature	0.07	экология ecology		0.13	Biodiversidad Biodiversity	Biodiversity	0.48		
2	Biodiversité Biodiversity	Biodiversity	0.07	Биоразнообразие Biodiversity		0.12	MedioAmbiente Environment	Environment	0.05		
3	biodiversite biodiversity	biodiversity	0.05	FAO		0.05	Naturaleza Nature	Nature	0.05		
4	Environnemer Environment	Environment	0.03	biodiversity biodiversity		0.04	naturaleza nature	nature	0.03		
5	Ecologie Ecology	Ecology	0.02	ЦПР Tsur		0.04	app	app	0.02		
6	environnemer environment	environment	0.02	UN	AND	0.04	PorLaNaturaleza	Corlantatura	0.02		
7	Chasse Hunt	Hunt	0.01	почва soil		0.03	España	Space	0.02		
8	PAC	PAC	0.01	природа Nature		0.03	Humedales	Wetlands	0.02		
9	Marseille	Marseille	0.01	ООН UN		0.03	medioambiente	environment	0.02		
10	animaux animals	animals	0.01	scientificrussia scientificrussia		0.03	conservación	conservación	0.01		

3. 結果と考察

3.1 ツイートの統計量

全ての言語コミュニティで投稿された Tweet の総数は 5,140,273 となった。最も多く呟かれた Tweet は英語 (2,925,742, 57%) とスペイン語 (1,422,724, 28%) であった。一方、ロシア語 (3,450, 0.7%) と中国語 (4,971, 1%) は少ない割合を占めた。その他、日本語 (301,571, 5.9%), アラビア語 (14,975, 0.3%), フランス語 (466,840, 9.1%) は 1 万から 50 万 Tweet であった。

3.2 定量・定性分析

全言語コミュニティで、Engagement は RT, like, reply で増加傾向が見られたが、前年に比べて減少する年も観測された (Fig. 1 a, b, c)。一方 diversity index は単調増加傾向が観測された (Fig.1 d)。全指標の動態から、Twitter は生物多様性に対する議論の場として活性化していることが示唆された。

Table 1 は全年度を通じて言語別に媒介中心性の高いハッシュタグを抽出した表である。日本語、アラビア語、中国語コミュニティでは組織やイベントが多く観測された。また英語、フランス語、ロシア語、スペイン語コミュニティでは自然や生物多様性を表す概念が媒介中心性の高いハッシュタグとして抽出された。言語コミュニティによって生物多様性に対する理解の違いが人々の生活に影響することが示唆された¹⁾

4. 今後の課題

今後、①value assessment でも求められている通り、多様な人々と協働することで伝統や文化を踏まえた深い生物多様性の概念を議論すること、②各言語コミュニティでの時系列的な話題の推移を可視化することを目指す。

参考文献

- 1) IPBES (2022). Methodological Assessment Report on the Diverse Values and Valuation of Nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Balvanera, P., Pascual, U., Christie, M., Baptiste, B., and González-Jiménez, D. (eds.). IPBES secretariat, Bonn, Germany. DOI: <https://doi.org/10.5281/zenodo.6522522>
- 2) Secretariat of the Convention on Biological Diversity (2020) Global Biodiversity Outlook 5. Montreal.
- 3) Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. Biological Conservation, 233, 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>
- 4) Tweepy (2022) <https://www.tweepy.org/>
- 5) NetworkX, (2022) <https://networkx.org/documentation/stable/install.html>