仮想的なタイムラインの再構築実験による エコーチェンバーの緩和についての検討

杉森 真樹 [1]　　笹原 和俊 [1, 2, 3]

[1] 名古屋大学大学院 情報学研究科
[2] 神戸大学 計算社会科学センター
[3]JST さきがけ

1　導入

現在の社会において, ソーシャルメディアは重要な役割を担うようになっており, 私たちの社会や生活に対して大きな影響力を持っている. このような背景のもと, ソーシャルメディアは議論の活性化やコミュニティの貢献に発展する一方で、実社会にネガティブな影響を及ぼしていることが大きな問題となっている.

その最たる例が「フェイクニュース」と呼ばれる虚偽情報の拡散とその結果として引き起こされる様々な社会問題である. 2016 年のアメリカ合衆国の大統領選挙に, ソーシャルメディアを介して, ロシアが関与していたとされる「ロシアゲート事件」を発端として, フェイクニュースが大きな注目を集めるようになった. フェイクニュースは, オンライン上で何らかの意図を持って拡散される虚偽情報であり, その拡散によって, 現実社会に様々な悪影響が生じることが懸念されている. このようなフェイクニュースが引き起こす問題の背景の一つには, 「エコーチェンバー」や「フィルターバブル」といった情報の多様性が失われてしまう不健全な情報環境の存在があると考えられている [7].

本研究では, 健全なソーシャルメディアのエコシステムの構築に貢献するために, ソーシャルメディアにおける不健全な情報環境の一つであるエコーチェンバーの緩和についての検討を行う.

2　エコーチェンバー

2.1　概要

社会においてソーシャルメディアが本来果たすべき役割は, 時間的または空間的な制約に縛られることなく, 様々な人間や情報が結びつくための機会を提供するプラットフォームとしての役割である. しかし, 近年ではソーシャルメディアが本来の役割を果たすのではなく, 社会的分断を助長してしまっていることが大きな問題となっている. これは, 人間が元来有する性質である, 同じ性質の人間同士で集団を形成するという「同類原理」や, 自分が信じたいものだけを信じる傾向があるという「確証バイアス」を強化してしまっていることが原因である. このような社会的分断が発生している状態は「エコーチェンバー」と呼ばれている [13].

2.2　関連研究

エコーチェンバーのような不健全な情報環境においては, 同様の意見を有するユーザーによってコミュニティが形成されていることに加え, 外部との繋がりが乏しいため, 同じような性質の情報が繰り返し拡散される. その結果, 情報がコミュニティ内で反響・増幅されることで, 自身とは異なる意見の情報に触れる機会が失われてしまう. このような情報環境の分離と同質化は, 集団的意思決定の阻害やフェイクニュースの拡散などを引き起こし, 実社会に悪影響を及ぼす危険性が指摘されている [11].

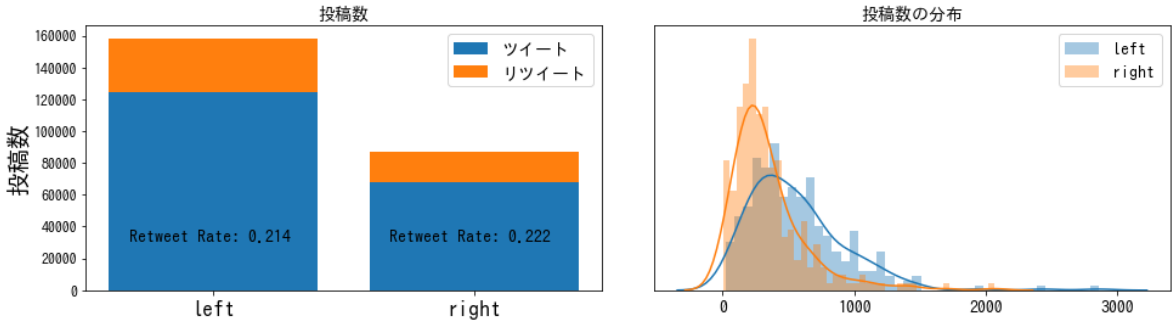

図 1: 党派ごとの投稿数と議員ごとの投稿数の分布

このような背景のもと, 健全なソーシャルメディアのエコシステムを構築するために, 不健全な情報環境の一種であるエコーチェンバーを対象とした様々な研究が行われている.

Adamic らは, まだソーシャルメディアが一般に広く普及していない, 2004 年のアメリカ合衆国の大統領選挙に関するブログのリンク構造において, Conover らは, 2010 年のアメリカ合衆国の中間選挙に関する Twitter の情報拡散において, リベラルな集団と保守的な集団に分離したコミュニティ構造の存在を確認している [4, 5]. Sasahara らは, オンライン上でのユーザーのアクティビティをエージェントベースモデルによって, シミュレーションすることで, エコーチェンバーの生成ダイナミクスの理解に取り組んでいる. また, エコーチェンバーを緩和するための原理を提案するとともに, その原理に基づいた独自のソーシャルメディアを開発し, 実験を行っている. [8, 9, 14].

2.3 本研究の取り組み

ソーシャルメディアにおけるエコーチェンバーに関連した研究は数多く行われているが, エコーチェンバーの程度を定量化するためのデファクトスタンダードとなる指標は存在しない. そのため, エコーチェンバーについて理解を深め, その緩和についての検討を行うためには, エコーチェンバーの程度を定量化するための指標を検討し, 定義する必要がある.

エコーチェンバーが引き起こす問題のひとつとして, 同じような情報が反響・増幅され, 異なる意見に触れる機会が失われてしまうという情報環境の同質化が挙げられる. これはユーザーが目にする情報の多様性が減少していると言い換えることができる. そのため, ユーザーが目にする情報の多様性に注目することで, エコーチェンバーの程度を定量化できると考えられる. そこで本研究では, ユーザーが閲覧する投稿に含まれる単語やハッシュタグ, ユーザーのエントロピーに注目する. そして, 後述するデータをもとに分析や実験を行い, エコーチェンバーを緩和するための検討を行う.

3 データ

3.1 データ収集

本研究では, 先行研究において社会的分断の発生が確認されているアメリカ合衆国の政治に注目する. アメリカ合衆国の政治は, 民主党（Democratic）と共和党（Republican）の 2 大政党が主導権を争っている. 民主党は, 「Left（左翼）」と見なされており, 平等な機会や権利などを重視し, 集団として社会を形成しようとするリベラルな思想の党派である. 共和党は, 「Right（右翼）」と見なされており, 軍事的な政策などを重視するとともに, 移民の受け入れに否定的など, 排他的で個人が独立した社会を形成しようとする保守的な思想の党派である. なお, 直感的な理解がしやすいように, 以降では, 民主党は「Left」, 共和党は「Right」と表記するものとする.

本研究では, アメリカ合衆国の政治に関するデータを収集するために, アメリカ合衆国議会の議員が発信する情報に注目する. 2019/4 - 2019/9 の 6 ヶ月間に渡り, 彼らの Twitter アカウントの投稿の収集を行った [3]. Twitter アカウントの存在が確認でき, 投稿が収集できたアカウントの数は, Left については 284 個, Right については 251 個であった.

3.2 データ分析

収集したデータについて基本的な分析を行う. 図 1 に, 全体の投稿数と各議員の投稿数の分布を示す. Left の方が, Right よりも投稿数が多く, 積極的に情報を発信していることが確認できる. リツイートについては, どちらも 20% ほどであり, 自ら作成した情報を投稿する傾向があることが確認できる.

次に, 集団としてどのようなトピックに注意を向けているのかを確認するために, ワードクラウドとハッシュタグクラウドの構築を行う. 図 2 にワードクラウドを, 図 3 にハッシュタグクラウドを示す.

ワードクラウドについては, Left と Right ともに一般的な単語が多く見受けられ, 大きな違いは確認できない. 一方で, ハッシュタグクラウドからは, Left と Right で利用しているハッシュタグに大きな違いが存在していることが確認できる. Left は「equalityact」や「lgbtq」などの平等な権利についてのハッシュタグが確認でき, 彼らの思想に沿った結果であることが確認できる. また, 「climatechange」などのハッシュタグの存在も確認でき, Left が Right よりも地球温暖化の防止にも積極的であることを反映している. Right は「usmca」などのアメリカ合衆国の政策に関するハッシュタグや「memorialday」などの軍事に関するハッシュタグが確認でき, 彼らの思想を反映していることが確認できる.

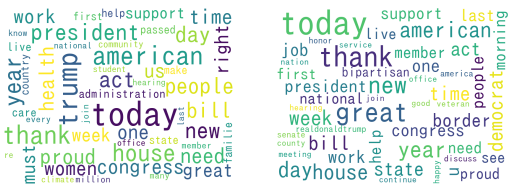

図 2: 議員の投稿に基づくワードクラウド

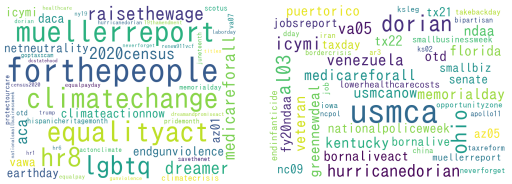

図 3: 議員の投稿に基づくハッシュタグクラウド

この結果から, 頻出単語からは党派による明確な違いは確認できないものの, 頻出ハッシュタグについては, 党派によって違いが存在しており, 注目しているトピックが異なっていることが確認できる.

本研究で利用するデータにおいて, コミュニティの分断が発生しているかの確認として, リツイートネットワークの構築を行う. 図 4 に構築したリツイートネットワークを示す. 青いノードは, Left のユーザーによってのみリツイートされたユーザーであり, 赤いノードは, Right のユーザーによってのみリツイートされたユーザーである. 紫のノードは, 双方の党派のユーザーからリツイートされたユーザーを表している. Left, Right それぞれがコミュニティを形成しており, 分断が発生していることが確認できる.

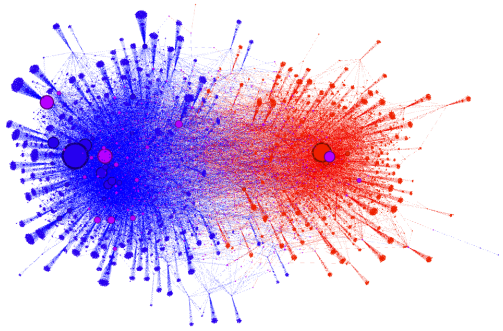

図 4: 議員の投稿のリツイートネットワーク

この結果から, 一方のコミュニティの情報のみに注目していると, そのユーザーが目にする情報は, そのコミュニティ内で拡散されている情報が大多数を占めることになると考えられ, 情報環境が分離, 同質化してしまっていることが確認できる.

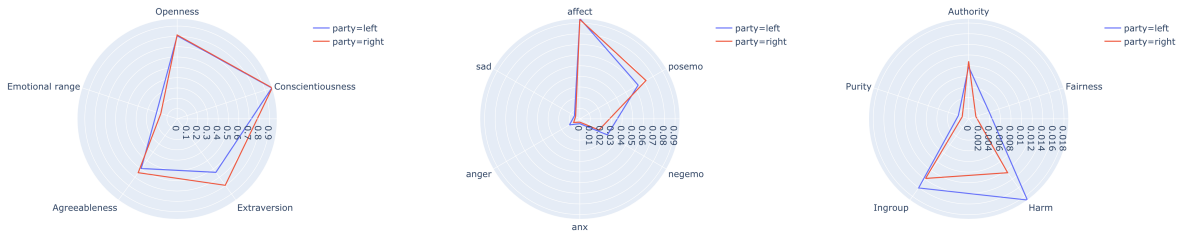

図 5: 党派単位の Big5, LIWC, MFD に関する値

4　言語的な特徴

図 3 のハッシュタグクラウドに反映されていたように, アメリカ合衆国の政治においては, 党派によって政策や思想などが異なっている. このような違いが実際にソーシャルメディアにおける投稿の内容に現れているのか, その違いがどのようなものであるかを確認するために以下の 3 つの側面に注目して分析を行う.

- 個人の性格（Big5）
  IBM Watson Personality Insights
- 投稿に含まれる感情
  Linguistic Inquiry and Word Count（LIWC）
- 投稿に含まれるモラル
  Moral Foundation Dictionary（MFD）

Personality Insights では, 投稿などの言語で表現された内容から, 人間の性格を「Big5」と呼ばれる心理学の分野で標準的に利用されている 5 つの評価尺度に加えて,「価値観」や「欲求」といった尺度によって評価することができる. なお本研究では、Big5 のみに注目するものとする.

LIWC は心理学の領域において, デファクトスタンダードとして利用されている辞書であり, 頻繁に利用される単語に品詞などの言語的なカテゴリーと感情などに関するカテゴリーを割り当てたものである [1, 10]. なお本研究では, 感情に関するカテゴリに注目して分析を行う.

MFD は,「道徳基盤理論（Moral Foundation Theory）」を様々なデータセットで検証するために作成された辞書である. 道徳基盤理論が仮定する「擁護（Harm）」,「公正（Fairness）」,「忠誠（Ingroup）」,「権威（Authority）」,「純潔（Purity）」の 5 つの道徳基盤と一般的な道徳に関連する英単語と語幹が登録されている [2, 6]. なお本研究では一般的な道徳を示すカテゴリは考慮しないものとする. また, モラル語自体の出現頻度が低いため, virtue と vice の合計の値を各カテゴリの出現頻度としている.

Left, Rigth の集団としての性質について分析を行うために, 各党派のユーザーの投稿をひとまとめにして, 党派単位の言語的な特徴の測定を行う. 図 5 に党派単位の Big5, LIWC, MFD の値を示す.

Big5 については,「Extraversion」以外の値に大きな違いは見られず, 集団としての Left, Right の性格的な特徴に大きな違いがないことが確認できる.

LIWC に基づく感情語が出現する割合については, 党派による差は確認できない. しかし, ポジティブかネガティブの内訳については, 党派によって異なっていることが確認できる.

MFD に基づくモラル語が出現する割合については, Left の方がより多くのモラル語を含んでいることが確認できる. この傾向は「Harm」,「Ingroup」,「Fairness」について顕著であり, 平等な権利などを重視する Left の思想が反映されている. 一方で,「Authority」については, わずかではあるが, Right の方が大きな値をとることが確認でき, 軍事などに重きを置く, Right の思想が反映されている.

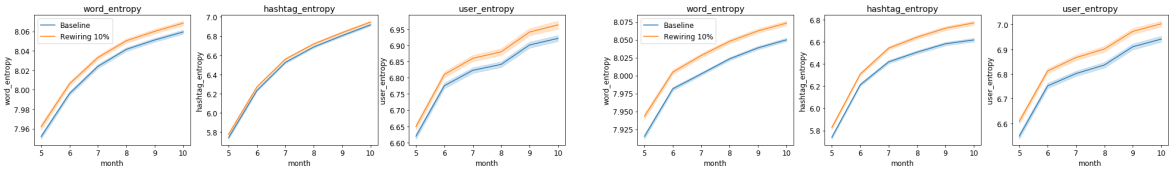

図 6: エントロピーの変化（左: Left -> Right, 右: Right -> Left）

5 実験

5.1 概要

エコーチェンバーが引き起こす大きな問題の一つとして, 情報環境の分離や同質化によって, ユーザーが目にする情報の多様性が減少することがあげられる. そのため, ユーザーが目にする情報の多様性を増加させることが, エコーチェンバーの緩和に繋がることが期待できる. 情報の多様性を増加させる方法の一つとしては, 情報環境に異質な繋がりを加えるために, ユーザー自身とは異なる意見を有するユーザーが発信する情報に注意を向けることがあげられる. そこで, エコーチェンバーを緩和するための検討の一環として, ユーザー自身とは異なる党派のユーザーをフォローすることで, ユーザーが目にする情報の多様性にどのような変化が生じるのかを測定するための実験を行う.

5.2 実験手法

情報の多様性がどのように変化するのかを測定するために, 収集した議員の投稿データをもとに, 仮想的なフォロー関係を作り, タイムラインを再構築し, 情報の多様性（エントロピー）の計算する実験を行う. なお, タイムラインとは, あるユーザーがフォローしているユーザー集合が行った投稿によって形成される投稿の集合体である. 実験の大まかな流れを以下に示す.

- いずれかの党派からユーザーを 150 人を選択.
- 任意の人数を無作為に選択し, 取り除く.
- 異なる党派から, 取り除いた数と同数を選択.
- タイムラインを再構築し, エントロピーを計算.

エントロピーは以下の 3 種類を測定する.

- 単語分布のエントロピー
- ハッシュタグ分布のエントロピー
- ユーザー分布のエントロピー

実験には, Left のみをフォローしている状態と Right のみをフォローしている状態からスタートする 2 通りのパターンが存在する. また, 各パターンについてのベースラインとして, フォローするユーザーを一切変更しないパターンの実験を実施する. 以降においては簡略化のために, フォローするユーザーを変更することを「Rewiring」と表現する. 1 回の実験において, ベースラインと Rewiring を行うパターンのそれぞれについて, 各 50 個ずつ仮想的なタイムラインを再構築し, 各種エントロピーの計算を行う.

また, 私たちの情報に対する認知のリソースの限界を考慮する必要がある [12]. そのため, ユーザーが目にする情報量が認知のリソースの限界を逸脱しないように, フォローするユーザー数を定める必要がある. そこで, 人間が安定的な社会関係を維持できる人数の上限とされているダンバー数に基づき, フォローするユーザー数を 150 人と定め, その 10% に相当する 15 人を Rewiring するユーザー数とする.

5.3 実験結果

図 6 に, 実験における各種エントロピーの変化を示す. 実線は 50 回の実験の平均をとった値を示し, 薄い色で示す領域は 95% 信頼区間を示すものとする.

Left の情報に注目している状態から Right のユーザーに目を向ける実験では, Rewiring によって, 単語のエントロピーとユーザーエントロピーが増加していることが確認できる. この結果は, ユーザー自身とは異なる意見のユーザーとつながることが, 情報の多様性の増加につながることを示している.

Right の情報のみに注目している状態から Left のユーザーに目を向ける実験については, 全てのエントロピーの値が増加していることが確認できる. これは, Rewiring の効果に加えて, 投稿数の増加が各種エントロピーの値に影響を及ぼしていると考えられる.

実験の結果から, 初期状態や集団の思想などの特性によらず, 自身とは異なる意見のユーザーとわずかでもつながることによって, そのユーザーが目にする情報の多様性が増加することを示した.

6 議論

ソーシャルメディアが引き起こす様々な問題の原因の一つであるエコーチェンバーの緩和についての検討として, アメリカ合衆国の政治に関するソーシャルメディアのデータ分析や, 仮想的なタイムラインを再構築し, 情報の多様性を測定する実験を行った.

その結果, 集団が注目しているトピックや情報の拡散過程についての分析することで, 社会的分断の発生を確認した. また, 実験によって, 自身とは異なる意見のユーザーと少量でも繋がることで, 情報の多様性が増加する傾向があることを特定した.

しかし, 私たちの情報に対する認知リソースは限られており, ユーザーは自身のタイムラインの投稿の内容を全て理解しているわけではない. そこで, 投稿時間に基づくフィルタリングを実装するなど, 認知リソースの限界を考慮した制約を加えた上で, 実験や分析を行うことで, 限られた認知リソースの中で, 情報の多様性を増加させるために目を向けるべき情報に関する知見の獲得が期待できる. その結果は, エコーチェンバーのような情報環境の分離, 同質化を緩和するための情報推薦システムの構築などへの応用が期待できる.

本研究で得られた結果をもとに, このような発展的な研究に取り組むことで, ユーザー同士のコミュニケーションを活性化するとともに, 多様な意見や考え方のユーザーが混在するコミュニティの構築や発展に貢献し, ソーシャルメディアのエコシステムの健全化に取り組む.

謝辞

本研究は, 科研費 JP19H04217 の助成を受けたものです.

参考文献

[1] LIWC. https://liwc.wpengine.com/.

[2] MFD. https://moralfoundations.org/.

[3] Us congress. https://www.congress.gov/.

[4] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: divided they blog. *International workshop on Link discovery*, 2005.

[5] Michael D Conover, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Partisan asymmetries in online political activity. *EPJ Data Science*, 2012.

[6] Jesse Graham, Jonathan Haidt, Sena P Koleva, Matt Motyl, Ravi Iyer, Sean Wojcik, and Peter H Ditto. Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 2012.

[7] Eli Pariser. *The Filter Bubble: What The Internet Is Hiding From You*. Penguin, 2011.

[8] Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. On the inevitability of online echo chambers. *arXiv*, 2019.

[9] Kazutoshi Sasahara, Giovanni Luca Ciampaglia, and Alessandro Flammini. エコーチェンバーの生成ダイナミクス. 人工知能学会全国大会, 2017.

[10] Yla R. Tausczik and James W. Pennebaker. The psychological meaning of words: Liwc and com-

puterized text analysis methods. *Language and Social Psychology*, 2010.

[11] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 2016.

[12] L. Weng, A. Flammini, A. Vespignani, and F. Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2012.

[13] 笹原 和俊. フェイクニュースを科学する. 化学同人, 2019.

[14] 笹原 和俊, Bruno Toshio Sugano, 奥田 慎平, and 佐治 礼仁. Polyphony 多様な繋がりを促進するsns. 人工知能学会全国大会, 2020.