令和7年度　インターネット上の偽・誤情報等への対策技術の開発・実証事業

偽・誤情報の拡散を抑制するためのSNSにおける シェア行動プロセス可視化と信頼性を評価する表示の検討

成果報告書 簡易版

2026/3/19

研04_東京大学大学院工学系研究科

偽・誤情報の拡散を抑制するためのSNSにおけるシェア行動プロセス可視化と信頼性を評価する表示の検討

アプローチする課題・目指す姿	• 偽・誤情報の流通・拡散の防止として、コンテンツモデレーションの手法が用いられているが、表現の自由との兼ね合いで透明性確保が難しいという限界があり、十分な偽・誤情報対策として機能しない可能性がある。このため、偽・誤情報の特徴を的確に捉える精緻な拡散行動メカニズムの解明が求められている。 • 信頼性情報の可視化によって、利用者の自律的な判断を促すメカニズムを解明することにより、偽・誤情報対策の迅速化と質的向上を目指す。		
研究・調査区分	偽・誤情報対策技術に係る研究	実施体制 (下線：研究・調査主体)	東京大学　鳥海研究室、(株)Lightblue、(株)電通

研究および有効性等に関する検証の取組・成果

- Chrome拡張機能の開発: Webブラウザ上で動作する拡張機能を開発し、Webページの構造解析（DOM監視）により投稿者情報を特定および、X（旧Twitter）上で、投稿者の過去のコミュニティノート（CN）付与履歴を可視化するツールを実装した。
- 実証実験による効果検証:募集した一般ユーザー38名を対象に、実際のSNS環境を模したシステムを用いて比較実験を実施。介入群（ツール使用）は対照群（ツール使用なし）に比べ、全体のリポスト数を約59%、CN付与投稿へのリポストを約73%抑制した。
- 長期的な行動変容: 追跡調査により、実験直後は高い抑制効果が見られるものの、2週間経過後には「慣れ」により効果が減衰する傾向（習慣化）を確認した。これらが報告書の主張と整合することを検証済み。

指標	対照群（ツール使用なし）	介入群（ツール使用あり）	抑制率
全体リポスト	7.37回	3.00回	**59.3%**
CN付与投稿へのリポスト	2.74回	0.74回	**73.1%**

研究・調査にあたっての課題・展望

- 大規模実証の必要性: 本研究では中程度の実用的な効果量（Cohen's d ≈ 0.5）を確認したが、統計的有意差の確立にはより大きなサンプルサイズが必要である。
- 習慣化への対策: 長期利用における効果減衰を防ぐため、UIの動的変更や、信頼性スコアに応じた介入強度の調整など、持続的な効果維持のための機能改善が求められる。

代表者コメント

東京大学大学院工学系研究科教授
鳥海不二夫

本研究では過去の信頼性の低い投稿の可視化が情報拡散の判断にどのような影響を与えるのかを明らかにした．投稿者が過去にコミュニティノートが付与された事実の可視化が，情報拡散抑制に効果的であることが示された．当研究の結果は偽誤情報対策に大きな貢献が見込まれる．