

第2回インフォメーション・ヘルスAWARD『社会実装部門』 応募用紙

○タイトル InfoCompass: 情報の海で迷わない、あなたの情報羅針盤

○応募者氏名

または グループ名 (株)TDAI Lab

○解決したい情報環境をめぐる課題

近年ソーシャルメディアの発展に伴い UGC が爆発的に増加する中で、個々のユーザーが視聴可能な情報量を遥かに超える状況(情報過多)が発生している。対策として多くの人によく閲覧されているもの(トレンドランキング)や、個人の過去の履歴などからレコメンド(おすすめ)されて視聴するという形態が広く浸透している一方で、偏ったものしか見られない(フィルターバブル)、同じようなものばかり見て考えが偏る(エコーチェンバー)などの影響を強く受けると、フェイクニュースに引っ掛かったり、社会的分断が起きることが指摘されている。

またユーザーはこれら懸念事項を理解していても、特定のトピック(e.g., 選挙)に関して、ネット上のニュースや世論において、現在どれほどの論点や意見が存在するか、また自身がどの程度それらを収集できているかを客観的に把握する方法がない。そのためユーザーは自身の情報収集環境外にある重要な情報や視点を欲したところで能動的に収集することが困難である。また推薦システムによって多様な視点を提供されたとしても、それによってどの程度新たな知見や視点を得ることができるのかが不明確であり、情報収集の効率や自身の情報収集行動を自省しようにも評価することが困難な状況が起きている。

○アイデアの具体的な内容(どんなもので、どんな人が、どう使うと、課題が解決できるのか)

我々は上記課題の根本的原因を「情報空間」の定量的把握することの困難さと可視化技術の不足だと考える。そこで我々は「InfoCompass」という情報空間可視化ツールを提案する。このツールでは、自然言語処理とデータ可視化技術を駆使して、特定トピックにおけるニュース、YouTube 動画、SNS コメントなどあらゆる情報ソースから情報を収集し、話されている論点の数や、それらへのコメントの参加量から「情報空間全体」の可視化を試みる。

ユーザーは本ツールを利用することで、情報空間全体の広さや意見の分布を俯瞰し、自身の情報収集状況を客観的に把握することが可能になる。具体的には、現在自分が全体のうち何パーセントの情報に触れているのか、またそれら情報がどの程度偏っているのかといった定量的フィードバックを行う。さらに同ツールは「メジャーな意見を知りたい場合」や「まだ見ていない重要な観点」といった情報集シナリオに応じて情報推薦を行う。そして、新たな情報を取得することで得られる「情報ゲイン」を可視化することで、多様な視点の獲得をゲーミフィケーション化し、意見の多様性・情報の信頼性の大切さに気付くきっかけを提示する。その結果、ユーザーがフェイクニュースなどに対して一定の「免疫」(批判的能力)を獲得することの実現を目的とする。

○アイデアは未発表のものかどうか。すでに「試作」「試行」している場合は、新たに付け加えたいアイデア(ブラッシュアップするポイント)など

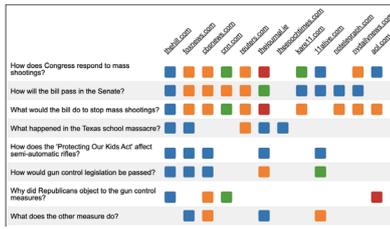
本アイデアの一部の機能については、すでに試作段階にある。我々は「東京都知事選」「総裁選」の2テーマについて、情報空間の可視化(各種候補者に対する賛否の投稿量・支持理由)に関する分析記事を複数寄稿しており、技術的に十分実用レベルであることが示されている。

本応募にあたっては、キュレーションアプリなどとの連携を意識し、「個人の情報消費履歴との連携」を強化し、より精緻で説明可能性が高い推薦を実現することを目指している。また、従来ニュースメディアなどとの連携も意識し、「リアルタイムでの情報空間更新機能」を追加することで、常に最新の情報環境を反映したマッピングを提供したいと考えています。さらに、ユーザーからのフィードバックを基に、可視化の精度と使いやすさを継続的に向上させていく予定である。

○アイデアを思いついたきっかけ

私は東京大学鳥海研究室にて博士を取得し、情報の多様性やバイアスの除去が研究テーマだったので、日常的にフェイクや炎上、極化といったソーシャルメディアに関する現象は身近な関心事でした。卒業後も NHK 様、鳥海研究室とは 2022 年以降『デジタル時代の放送における正確な情報発信に向けて』といったテーマで共同研究も行なっており、日常的に如何にネット上の多様性を測定するかや、どうすればユーザーがネット上の偏りや誤情報に気づいてくれるかの情報収集、研究活動を行なっています。

そして以前読んだ論文に「Discord Questions: A Computational Approach To Diversity Analysis in News Coverage」といった論文があり、ここではキュレーションしたニュースに対して、それぞれの記事がどの観点を含んでいて含んでいないかといった可視化を行ない、それを見せてユーザーに情報の多様性に関する意識調査を行うといったものでした。(以下図) これを見た際に自分たちもネットコメントに対して同じような視点で分析を行いたいと感じました。しかしネットコメントの量の多さから同様の可視化は不可能であり、その経験が効率的な情報空間の可視化について考えさせられるきっかけとなりました。



○アイデアを実現するための技術的裏付けや学術的裏付け

「情報空間の可視化」実装には「広範なデータ収集」「LLM を用いた構造的情報収集」「クラスタリング」の 3 種類が必要になる。(補足資料参照)。

「広範なデータ収集」に関して：我々は業務の一環で、独自の情報収集クローラーを開発しており、X の投稿収集、YouTube の関連動画収集、コメント収集、動画文字情報収集など収集機能を実装している。また特定テーマに対しては 1 日 2 回収集するといったリアルタイム性を伴う収集技術も開発済みである。

「LLM を用いた構造的情報収集」：我々の情報空間可視化の独自性は(entity(人物・団体など), stance(賛否、期待、悲観、各種感情など), reason(理由))といった triplet 構造の収集によって達成される。我々は広範なサーベイ・仮説検証の元、これらに基づいた集計が最も説明性や、分析結果としての納得度が高いと今のところ判断している。コメントといったノージーな非構造データからこれら triplet 構造抽出を行う過程において LLM を活用している。我々は現在プロンプト管理のテンプレートと、全コメントに対して 1 件 1 件行うことができる並列処理の仕組みを実装しており、高速な PDCA が可能になっている。

「クラスタリング」：上記 triplet から抽出した情報について、embedding を作成し、HDBSCAN を用いることで、entity の名寄せや reason のユニークな個数を見つけることができる。これら手法選択も最新の研究をベースに選んでいる。

繰り返しになるが、今後の実装方向性は、「個人の情報消費履歴との連携」「リアルタイムでの情報空間更新機能」「ユーザーからのフィードバック」を基にした可視化の精度と使いやすさを継続的に向上がテーマになる。これらは高い実装能力と Human-Computer Interaction 分野に関する深い知見が必要になり、学術的知見と高い実装能力をもった東京大学鳥海研究室発 AI ベンチャーである弊社ならではの遂行できると確認している。

○アイデアを実現するための期間および資金、具体的な工程表

InfoCompass の実現には、計 2 年の段階的なアプローチを取る予定です。

第 1 フェーズ「情報可視化機能のブラッシュアップ」(3～6ヶ月)：まずは従来新聞・テレビメディアのジャーナリストなどに協力を仰ぎ、ネット上のあらゆる事象について情報空間の可視化、ネット上の偏りを可視化する機能を見ていただき、世の中に出せるレベルになるまであらゆるケースでフィードバックをいただく。これは分析結果と分析記事になるまでには大きなギャップがあるという弊社のこれまでの知見によるものである。そしてこの過程で triplet の見直しや、追加回収項目の洗い出しを行う。

第 2 フェーズ「リアルタイム自動作成機能の実装」(3ヶ月)：続いて要件が決まったのちに、特定のテーマだけ入力すれば自動で第 1 フェーズの成果物が出るまで実装を自動化する。その過程でジャーナリスト等が修正を加えたい場合、再生成などを可能にする Human-in-the-Loop の仕組みも搭載し、より実用的なものにする。この時点で各メディア機関などがデジタル世論分析などが社内で使えるプロダクトが完成される。

第 3 フェーズ：「個人データとの統合」(3～6ヶ月)：各種ベンダーと協力し、ユーザーの視聴データから全体の情報空間に占める情報取得状況のフィードバックをリアルタイムで可視化できるようにする。その過程で推薦技術も実装する。

第 4 フェーズ：「ユーザー受け入れフェーズ」(6ヶ月)：ユーザーテストの結果を反映しつつ改善と、大規模データ処理への対応を進める。最終的に、これらのプロセスを経て正式リリースを行い、その後も継続的な改善を行っていく予定です。

資金に関しては、開発規模や採用する技術によって大きく変動する可能性があります。5 千万円程度の投資が必要になると見込んでいます。

○アイデアを実現するための実施体制・能力（専門的スキル、知識や経験など）があれば付記

TDAI Lab は 2016 年創業の東京大学鳥海研究室発 AI ベンチャーです。

代表福馬は 2022 年に同研究室にて博士号取得。研究領域はネット上のバイアス・公平性・推薦など。

著書に「世界一カンタンで実践的な文系のための人工知能の教科書」

TDAI Lab は、最新の AI 技術の企業への導入支援を行っており、要件定義、データ収集、PoC、システム開発、スキルトランスファーなど AI 開発に関わる全工程を一社内製で行っており一貫した開発を得意とする。

2022 年以降 NHK 様、鳥海研究室と『デジタル時代の放送における正確な情報発信に向けて』といったテーマで共同研究も行なっており、情報空間分析に関するシステム開発・技術開発・論文執筆なども手掛けています。

今回のテーマにおいて、これらネット情報空間の偏りに関する深い学術知見と、それを実装まで持っていく企業としての実装能力の両方の観点から、本募集の社会実装部門としては最適であると自負しています。

インフォメーション・ヘルス AWARD2024
社会実装部門補足資料

“InfoCompass: 情報の海で迷わない、あなたの情報羅針盤 ”

2024年9月30日

株式会社TDAI Lab



デジタル時代における情報偏重の課題と提案ソリューション概要

現状: 情報過多と偏りの発生

- **情報過多の加速** : デジタルコンテンツの爆発的増加により、個々のユーザーが処理できる情報量を遥かに超える状況が発生
- **情報の偏在化** : 「いいね」推薦アルゴリズムにより、一部コンテンツや既存の興味に基づいた情報が優先的に表示される傾向が顕著化
- **技術的対応策** : 多様性のある推薦やLLMを活用した要約が、多様な視点の提供を目的として実装されている

ユーザー体験の課題: 情報収集の不透明性

- **情報空間全体の不可視性** : 特定トピックにおいて現在全部でどれだけの意見があるか、また自身がどれだけ収集できているかが**定量・客観的に** 把握できない
- **ブラインドスポットの存在** : 自身の**情報環境外**にある重要な情報や視点を認識できない
- **新規性の不明確さ** : レコメンドなどによって、情報を提供されてもそれによって**どれだけ自分が新規性**を手に入れられるのか把握することが困難

提案ソリューション: 情報空間全体の可視化

- **情報空間の可視化技術の開発** : 自然言語処理とデータ可視化技術を活用し、あらゆる情報ソース(ニュース・動画・コメントなど)から、「**情報空間全体**」を可視化
- **獲得情報量と情報多様性のスコア化** : 現在自分が全体のうち、何%の情報に触れているのかわかる。またそれがどれほど偏っているかを定量化する
- **パーソナライズされた推薦** : 「メジャーな意見を知りたい場合」や「まだ見ていない重要な観点」などシナリオに応じた推薦、またそれによって得られる情報ゲインを伝えることで多様な視点獲得をゲーミフィケーション化

自分の知らない情報が
どれだけあるかがわからない

デジタル時代における情報偏重の課題と提案ソリューション概要

現状: 情報過多と偏りの発生

- 情報過多の加速
り、個々のユーザー
状況が発生
- 情報の偏在化:
コンテンツや既存の
れる傾向が顕著
- 技術的対応策:
が、多様な視点の

ユーザー体験の課題: 情報収集の不透明性

プラットフォームにおいて現在全部で
どれだけ収集できてい

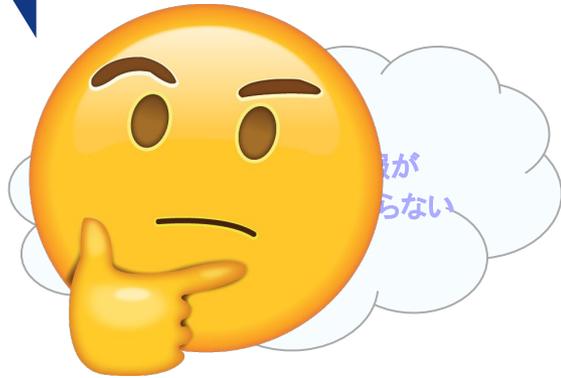
情報環境外にある重要

って、情報を提供さ
規性を手に入れられ

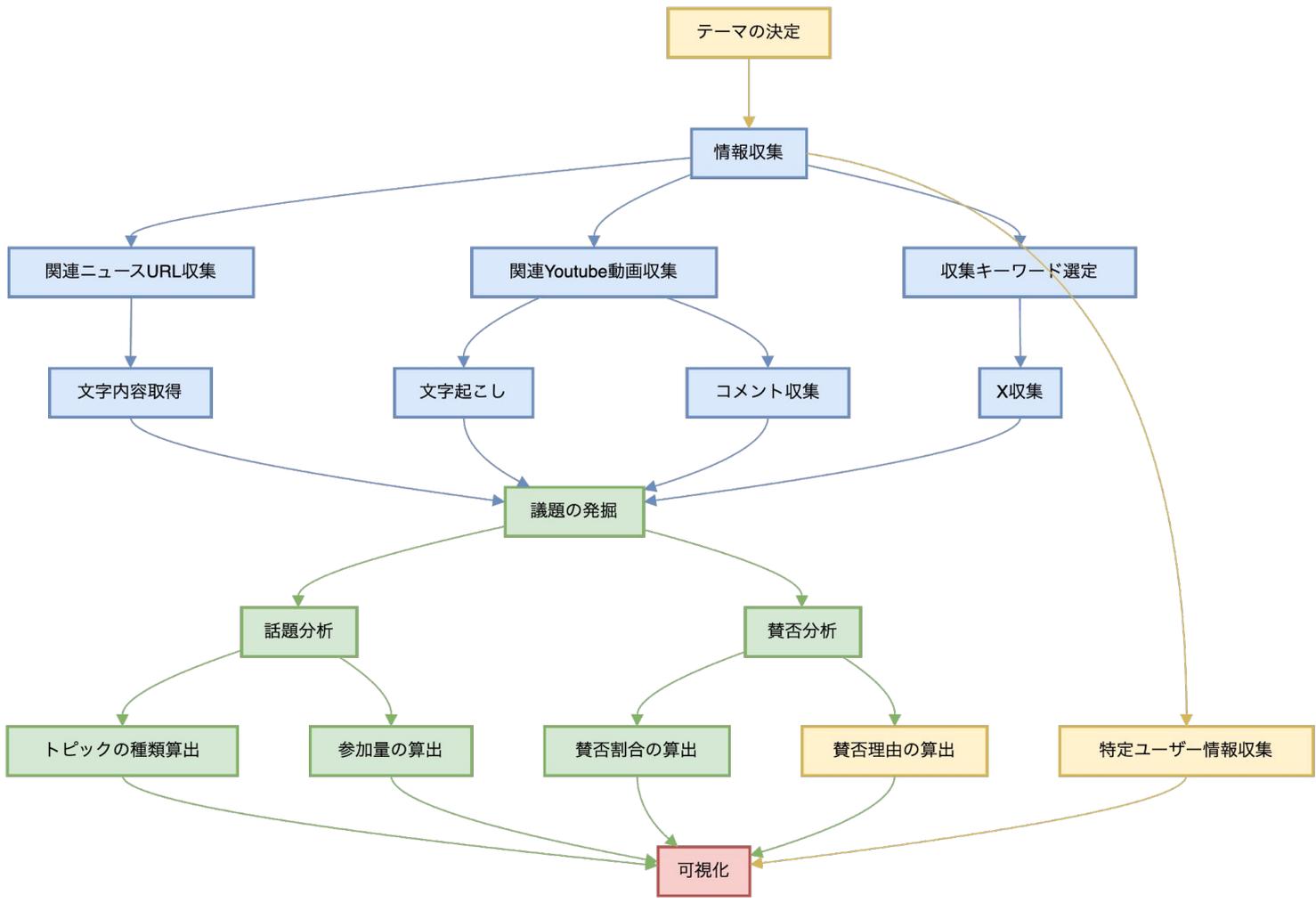
しかし情報空間の可視化はどうやって行う？

提案ソリューション: 情報空間全体の可視化

- 情報空間の可視化技術の開発 : 自然言語処理とデータ可視化技術を活用し、あらゆる情報ソース(ニュース・動画・コメントなど)から、「情報空間全体」を可視化
- ↓
- 獲得情報量と情報多様性のスコア化 : 現在自分が全体のうち、何%の情報に触れているのかわかる。またそれがどれほど偏っているかを定量化する
 - パーソナライズされた推薦 : 「メジャーな意見を知りたい場合」や「まだ見ていない重要な観点」などシナリオに応じた推薦、またそれによって得られる情報ゲインを伝えることで多様な視点獲得をゲーミフィケーション化



可視化までの実装の流れ



データ収集について

特定のテーマに対して幅広いプラットフォームから関連情報を取得



Google Newsからは各種メディアの記事を収集



Youtubeからは動画内容とコメント情報を取得



Xからはユーザー投稿を

情報空間の可視化に求められる機能的要件

機能1.論点の数の把握

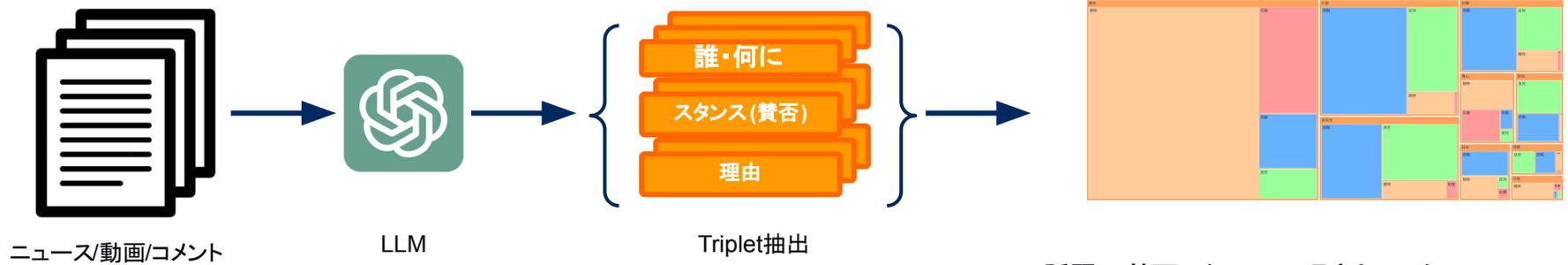
メディア(ニュース、関連動画、コメント等)から論点(話題・意見)を、LLMを用いた抽出とクラスタリングにより、**全ての論点を抜け漏れなく把握**

機能2.論点の偏り把握

動画やコメントが各話題に対してどれだけ参加しているかを把握。それによって情報空間上でどれだけ一部の情報が話されているかを把握

3.個人の摂取状況の把握

個人の閲覧ログなどから、どの論点ハマりにして、どの情報は見ていないかを把握し、情報接種の網羅性や公平性を定量評価



- ・話題 → 賛否スタンス → 理由といった階層構造での可視化を想定
- ・面積が量を表す
- ・色は将来的に個人の収集達成状況を表す

類似実績：総裁選前コメントを用いた情報空間の可視化

- 9/12 - 9/26「総裁選」に関する動画を再生回数 200本から100コメントずつ収集
- 1万件から各候補者へのスタンス（期待・応援・悲観・反対）と理由を gpt-4-o-mini を用いて収集



	コメント数	理由数	%
高市	1589	358	22.530
石破	259	132	50.965
小泉	556	209	37.590

それぞれの候補者についてコメント数と理由の集計
 (高市氏に関するコメントは数に対して理由の多様性が少ない)

ネット上の意見は高市氏に対する期待の声、小泉石破氏に対する否定的な声が顕著

ツールの社会実装に向けて

- **第1フェーズ「情報可視化機能の強化」(3～6ヶ月)**
 - ジャーナリストに協力を仰ぎ、ネット上の情報空間や偏りを可視化する機能をテスト。
 - フィードバックを基に改良。
 - tripletの見直しや追加項目の精査を実施。
- **第2フェーズ「リアルタイム自動作成機能の実装」(3ヶ月):**
 - 特定テーマを入力すると自動で第1フェーズの成果物が出るように実装。
 - Human-in-the-Loop機能を追加し、ジャーナリストが修正できる仕組みを導入。
- **第3フェーズ「個人データとの統合と推薦・ゲーミフィケーションの考察」(3～6ヶ月):**
 - ベンダーと協力し、ユーザーの視聴データを基に情報取得状況をリアルタイムで可視化。
 - 推薦技術の導入により、ユーザーに最適化された情報提示を実現。
 - ゲーミフィケーションの要素を考慮し、ユーザーが楽しく情報を獲得できる仕組みを検討。
- **第4フェーズ「ユーザー受け入れフェーズ」(6ヶ月):**
 - ユーザーテストを基に機能を改善し、大規模データ対応を進める。
 - 正式リリース後も継続的に改善。