

実社会 SNS 情報を用いた LLM エージェントベース SNS シミュレーション

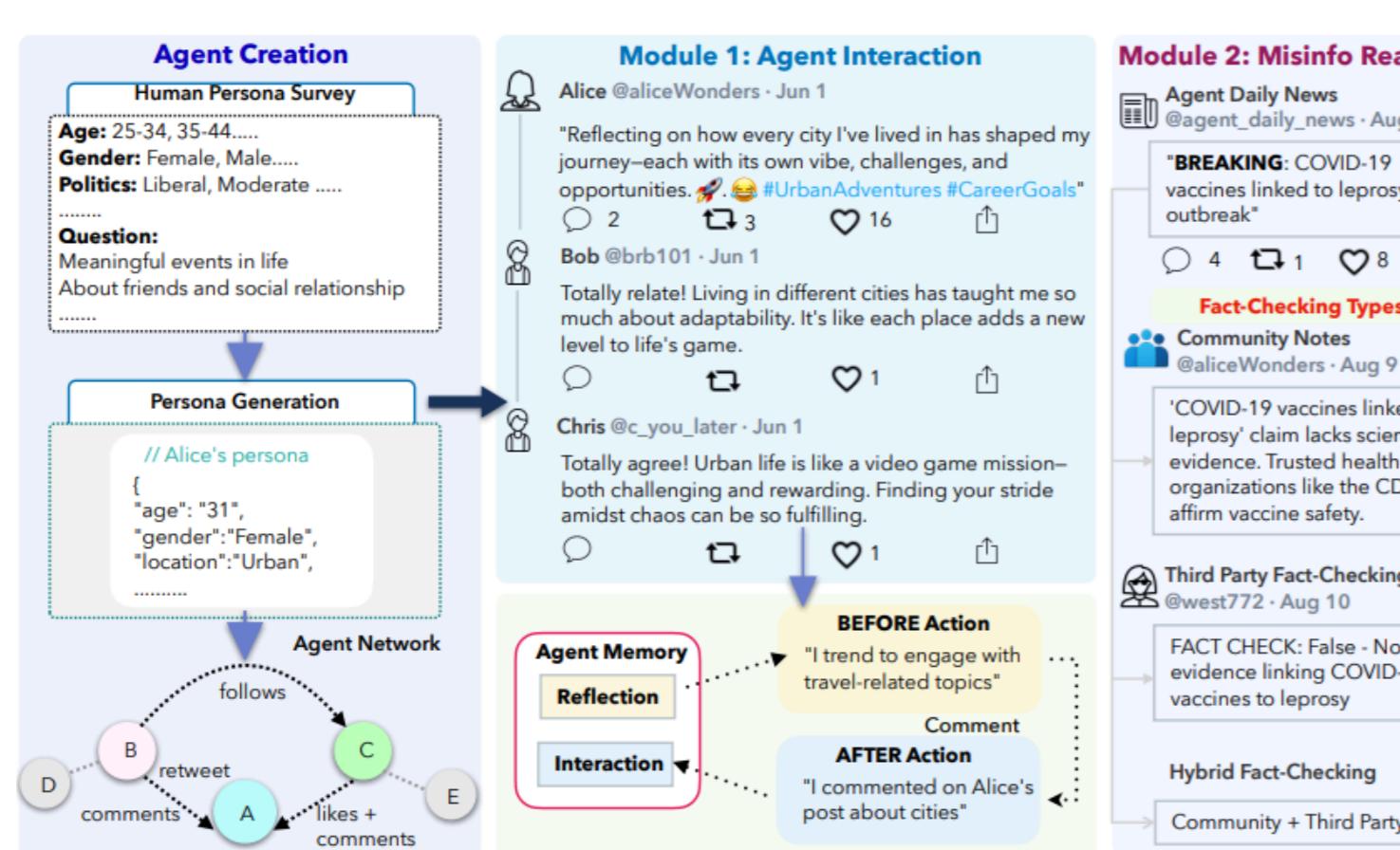
村上研究室 f21038 5J11 小山弘人

背景・目的

SNS 上での情報拡散や世論形成のメカニズムを理解するためには、実社会の SNS データを用いたシミュレーションが有効である。しかし、既存の研究では実社会の複雑なユーザー行動や相互作用を十分に再現できていない。本研究では LLM エージェントを活用し、実社会 SNS データに基づくシミュレーション手法を目的とする。

先行研究

先行研究では、被験者 204 名のアンケートから得た人格データを用いて LLM エージェントをモデル化し、シミュレーション環境を構築している。この環境で、ユーザー行動の予測、誤情報の拡散過程、およびファクトチェック戦略の影響を分析している。



シミュレーション構成

初期人数:50 人

タイムステップ:40 回

LLM:gpt-4o, claude-3.7-sonet, deepseek-v3

ファクトチェック:第三者, コミュニティノート, ハイブリッド型

ステップごとの処理は以下のようになっている。

- ニュースを投入する
- ランダムで新規ユーザーを投入する(最大 2 人)
- フィード内にあるニュースに対する反応を LLM に基づき生成
- 定期的に自分の行動を記録する

このシミュレーションで、誤情報は正しい情報と比べて拡散しにくいことや、使用した LLM やファクトチェック戦略によって効果の違いがあり、結果 claude-3.7-sonet のハイブリッド型が最も効果的であることが示された。

それに加え、シミュレーションの再現度を検証するため、被験者 204 名と対応するエージェントに対し、誤情報を含む 30 件の投稿を提示する比較実験を行った。得られた反応の平均値を独立二標本 t 検定により分析し、両者の行動パターンに統計的な整合性があることが示された。

現在の進捗と今後の展望

現状、Bluesky でのデータ収集と並行してシミュレーションの実装を進めている。Bluesky のユーザーは現在 300 万ユーザー分の収集を完了しており、最終的には 3000 万ユーザー分の収集を目指している。またシミュレーションは先行研究のコードをベースに実装しており、現在はペルソナ生成のためのコーディングをしている段階である。

今後の展望としては、まずシミュレーションの各種設定を実社会データに基づき調整・実装し、最終的にはより実社会データに基づくシミュレーション環境にてコミュニティの成長や話題の発生や拡散を分析する予定である。

提案手法

先行研究の手法は比較的正確なペルソナの再現が可能であるものの、抽象的な属性データへの依存度が高く、SNS 特有のコミュニティ構造や文脈の再現には限界がある。加えて、アンケート調査には多大なコストを要するため実験規模は数百人程度に留まっていた。そこで本研究では、実社会の SNS データを用いることでこれらの課題を解決し、大規模かつ現実的なシミュレーション環境を構築することを提案する。

1.ペルソナの生成

まず、SNS データを Bluesky API を用いて収集し、ペルソナ生成に必要なユーザー属性や行動データを抽出する。

ペルソナ生成に用いるデータは主に以下の通りである。

- ユーザーのプロフィール説明文
- 投稿テキスト
- リツイート投稿テキスト
- フォローアクションのプロフィール説明文や投稿テキスト

これらを用いてユーザーの興味関心や行動傾向を LLM で分析し、ペルソナを生成する。

2.投稿頻度のシミュレーション

次に、ユーザーごとの投稿頻度を設定する。先行研究ではタイムステップ単位で同時に投稿や反応が行われていたが、実社会ではユーザーごとに投稿頻度が大きく異なる。そこで、収集したデータから各ユーザーの平均投稿頻度を割合で算出し、シミュレーション内でその頻度に基づき投稿行動を決定する。これにより、より現実的な投稿パターンを再現することが可能となる。

3.フォローネットワークの評価

先行研究では、フォローネットワークの初期化に以下のようないわば Barabási-Albert モデル(以降 BA モデル)を採用していた。

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^N k_j}$$

BA モデルはスケールフリー性が理論的に保証されるため、構造の検証を行わなかった。本研究ではネットワークを新規に構築するため、その妥当性の検証が必要となる。そこで、Bluesky の収集したフォロー関係と比較し、現実の SNS 特性が再現されているかを評価する。

4.フィードアルゴリズムの設定

最後に、フィードに表示するポストのアルゴリズムを設定する。先行研究では時系列順に投稿が表示されていたが、より SNS に近いメカニズムを再現するためには協調フィルタリングなどの手法を導入する必要がある。そこで、収集したデータを基に、ユーザーのフォロワーなどの属性が近いユーザーがリアクションを行ったポストを優先的に表示するアルゴリズムを実装する。