

はじめに

概要

まとめ

## 【研究紹介】 ソーシャルネットワークにおけるフィルターバブルに関するシミュレーションの研究

武藤 克弥 (Katsuya Mutoh)  
u255018@st.pu-toyama.ac.jp

富山県立大学 大学院 電子・情報工学専攻 情報基盤工学部門

April 28, 2023

## 背景

- WEB の推薦アルゴリズムによって、情報の孤立・偏りが生じるフィルターバブルが問題視されている.
- アルゴリズムを改善し、検証する環境が十分に整っていない.

## 目的

- SNS における記事の閲覧行動をモデル化し、マルチエージェントシミュレーションによって、推薦アルゴリズムをテストできるようにする.
- フィルターバブルを起こさず、かつユーザに有益な情報を提供できる方法を提案.

## フィルターバブル現象と SNS に求められること

- WEB や SNS ではユーザの好みに合わせた情報が提供されやすい.
- 自分の好みに合った情報・似通った意見の人が周りに集まり, 偏った情報が正しいと錯覚してしまう.
- フィルターバブルを防ぐと同時に, ユーザごとに有益な情報を提供し続ける必要.

## 改善アルゴリズムの評価

- SNS 運営が改善した推薦アルゴリズムを検証するのは難しい (多量のデータ+社会的影響).
- 検証も可能な SNS モデルを作成し, フィルターバブルを改善するアルゴリズムを評価可能にする.

## 前提条件

- ユーザエージェント  $N$  人.  $N$  人は相互にフォロー
- ユーザは 0~99 のクラスタ (100 種) のいずれかに所属 (≡趣味)
- 記事エージェント (0~99 のクラスタ)
- クラスタ値が近い=カテゴリが近い
- 平均 (ユーザのクラスタ値), 標準偏差 0.5 とする正規分布で記事のクラスタが決定
- 記事閲覧ごとにユーザのクラスタは変化する (確率)

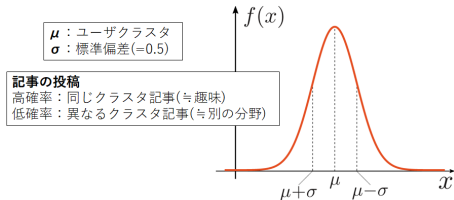


図 1: 記事エージェントのクラスタ割り当て

## 前提条件 2(ユーザの行動)

- ステップ数=投稿記事の時系列
- 1 ステップ= $N$  人から 1 ユーザを選択  
→ 1 人が記事の投稿 or 閲覧を行う
- ユーザはニュースフィード (タイムライン) を閲覧  
→ 確率でエンゲージメント (いいね RT) を与える.

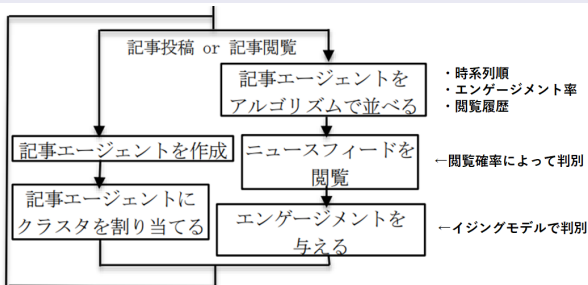


図 2: 1 ステップの流れ

## 記事の閲覧

- ① アルゴリズム (後述) で記事を並び替え
- ② 確率でユーザは記事詳細を閲覧 (MAX30 記事までしか見ない)
- ③ ユーザはニュースフィード (タイムライン) を閲覧  
→確率でエンゲージメント (いいね RT) を与える.

②記事詳細の閲覧確率  
=TS × CS

タイムラインスコア(TS):

1~0(上の記事ほど高い)

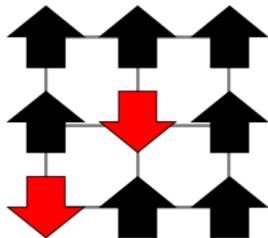
ユーザ・記事クラスタ類似度(CS):

$$1 - \frac{(\text{ユーザ・記事クラスタ値の差})}{100}$$

③エンゲージメントする確率=イジングモデルで判定

## イジングモデル

- ① 1つのスピンの周りに4つのスピン (格子状に配置)
- ② +1 か-1 のいずれかの値 (周りの多い方に遷移しやすい)
- ③ エネルギー  $E$  が小さくなるようにスピンは状態遷移
- ④ 確率  $P$  でエンゲージメントを与える



初期状態:  $S_0 = -1$   
 記事投稿:  $-1 \rightarrow +1$   
 $+1$ で閲覧: いいね  
 $-1 \rightarrow +1$ 閲覧時(確率 $P$ で): RT

$$E = - \sum_{i=1}^4 J \times S_0 \times S_i$$

$E$ : エネルギー

$S_0$ : 自分自身のスピン

$S_i$ : フォロワーのスピン

$$T = \frac{1}{1 + 2 \left| \sum_{j=1}^4 S_{ij} \right|}$$

$J = 0.047$ : 磁性強度

( $\equiv$  人間関係の強度)

$T$ : 温度

パラメータ

$$P = \exp\left(-\frac{\Delta E}{T}\right)$$

$P$ : 状態遷移 (記事に反応する) 確率

$\Delta E = E$  更新後 - 更新前

$$M_j = \frac{1}{N} \sum_{i=1}^n S_i$$

$M_j$ : 記事  $j$  の情勢 (注目度)

$\leftarrow 0$ に近い: 安定

1に近い: バズリ・炎上

タイムラインに表示される記事のソート基準

## 時系列アルゴリズム

- 記事を新しいものの順に並べタイムラインに表示
- 自分に価値のない情報が多くなる

## エンゲージメント率アルゴリズム

- いいね RT 数を閲覧数で割った割合 + 時系列スコア でソート
- バズりの記事に関連するものが多くなり、自分に価値のない情報に偏る

$$Eg = \frac{5 \times \text{RT 数} + \text{いいね数}}{\text{閲覧数}}$$

## 閲覧履歴アルゴリズム

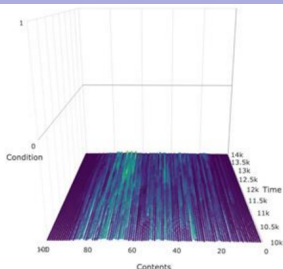
- 過去にいいね RT した記事と同じクラスタで加算 + 時系列スコア でソート
- 自分に価値のある情報が並ぶが、情報が偏る



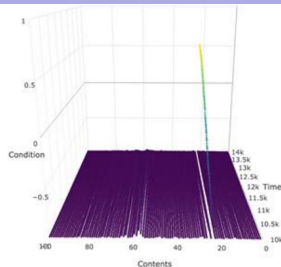
従来アルゴリズムを用いて、モデルが妥当であるかシミュレーション

## 従来アルゴリズムによるシミュレーション

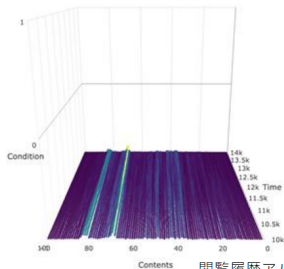
- ① ユーザエージェント数:1225 人
- ② ステップ数：4000 回
- ③ あらかじめ 10000 ステップ実行 (記事の数を十分に作る)



時系列アルゴリズム



エンゲージメント率



閲覧履歴アルゴリズム

縦：ステップ数(1万～1.4万)  
横：記事クラスタ  
高さ：記事の情勢M

従来アルゴリズムを提案したモデルで再現

図 3: 各アルゴリズムでのモデルシミュレーション

フィルターバブル (情報の偏り) を起こさない, かつ有益な情報提供ができるもの

## 従来アルゴリズムの長短

- 時系列情報の偏りが無いが, 自分の嗜好に近い情報が少ない
- エンゲージメント率少し価値があるが, 情報の偏りが激しい (同調圧力)
- 閲覧履歴自分の嗜好に近いが, 情報が偏る

## フィルターバブルの数値化

- タイムラインに表示される記事クラスタの偏りを算出
- クラスタ値の標準偏差を使用
- 全ステップ全ユーザの平均をとる

## 有益な情報の数値化

- 不便指数と定義 (低いほど良い)
- タイムライン上位 30 個の記事に対し，ユーザクラスタ値との絶対値の差を平均化
- 全ステップ全ユーザの平均をとる

1ステップ1ユーザにおける不便指数(興味のない記事の表示度合い)

$$\text{不便指数} = \frac{\sum_{k=1}^{30} |(\text{記事 } k \text{ のクラスタ値}) - (\text{ユーザ } i \text{ のクラスタ値})|}{30}$$

## 評価

- ユーザーエージェント 1225 人, ステップ数 4000 で再実行
- 10 回施行した平均を出力

Table 1: 各アルゴリズムでの評価

	TIME	ENGAGE	HISTORY
標準偏差	18.95	3.94	7.01
不便指数	19.50	16.03	12.51

- 標準偏差を時系列アルゴリズムに近づけ, 不便指数を閲覧履歴アルゴリズムに近づけるが理想

→ 閲覧履歴をもとに自分の嗜好にあったものを表示  
+ 時折嗜好ではないものを混ぜるのが妥当

## 提案手法のシミュレーション

提案手法=閲覧履歴をもとにした自分の嗜好とは異なるものをタイムラインに混ぜる

- どれくらい混ぜれば妥当か，3%ずつ増やして検証 (0~60%)
- タイムライン内の標準偏差を不便指数を算出
- ユーザエージェント 1225 人，ステップ数 4000 で実行
- 10 回施行した平均を出力

## 実験結果

- 標準偏差 9%以上で時系列に近づく．以下で閲覧履歴に近づく
- 不便指数 9%以下で閲覧履歴，以上で時系列に近づく

→ 9%ユーザ嗜好以外の記事を混ぜる=フィルターバブル回避+有益な情報提示

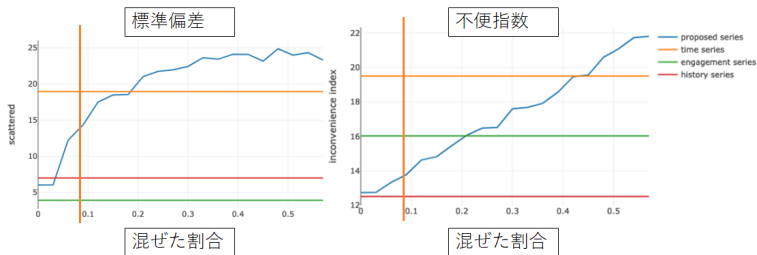


図 4: 提案手法の評価

## まとめ

- 推薦アルゴリズムを検証する SNS モデルとシミュレーション環境を構築した
- フィルターバブルをなるべく防ぎ、より有益な情報をユーザに提供する方法を提案

## 今後の課題

- 実際の SNS で試験運用してさらなる検証が必要  
(実際の SNS (ユーザ数・ステップ数が未知数) に対する評価)