

ツイッターコンテンツから生成される情報の偏りに関する研究

陸 易[†] 高橋 幸雄[†] 一島 力男^{††}
[†] 国士舘大学理工学部 ^{††} 国士舘大学体育学部

1. はじめに

情報の繭とは、「我々が住むコミュニケーションの世界で我々の好みしか現れない世界」であり、キャス・サンステーンは現代人が「情報の繭」の中に囚われている指摘した[1]。フィルターバブルとは、推薦などの技術によって、情報の偏りが利用者の知らぬうちに発生するという問題である[2]。上記二種類の問題は情報の偏りにより構成される問題。本研究は Twitter から取得したデータからレコメンドアルゴリズムを分析し、情報の偏りの生成過程を研究することを目的とした。

2. 提案手法

2.1 レコメンドアルゴリズム誘導

テスト用のアカウントを複数準備し、『Z 世代が選ぶ 2022 上半期トレンドランキング』サイトから、検索のキーワードを選んだ。Twitter でキーワードを検索し、行った操作を全て記録した。

2.2 ホームページタイムラインデータ取得

各テストアカウントでログインし、ユーザーホームページのタイムラインデータを取得した。完了後、実験記録テキストファイルを開き、テーマやレコメンドツイートか否かを記録した。

2.3 データの視覚化

得られたツイート文データを分類して、出現率表とグラフを作った。

3. 評価実験

次の二段階の実験を行った。第一段階は 8 つのアカウントが同じコンテンツを目的とした場合、「キーワード検索」「アカウントフォロー」「いいね・リツイート」をコントロールして実験した。その結果アカウントに表示される影響はアカウントフォロー > いいね・リツイート > キーワード検索となった。第二段階では、1つと 3 つのアカウントの二組が二種類のコンテンツを目的とした場合、アカウント間の差を見出す実験である。2組の三アカウントは「フォローしたアカウント」「いいね・リツイート」をコントロールした。

図1は一例として、Twitter レコメンドアルゴリズムよりおすすめされたツイートが、ユーザーのホームページタイムラインで表示された割合を示したものである。第一段階の実験では、レコメンドツイートの割合は 70% 上下で浮動していて、第二段階では平均 55% までに下がった。つまり、アカウントの使用頻度の上昇とともに、レコメンド

に関するコンテンツはある程度減少する。

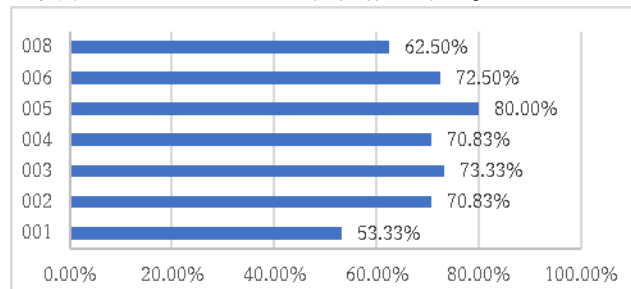


図 1:レコメンド割合の一例

もう一つの評価指標はピアソンの相関係数を用いた。ピアソンの相関係数とは 2 つの変数の間の類似性の度合いを指す統計的指標である[3]。表1に示した通り、第二段階のデータで各アカウント間のピアソン相関係数を算出した。アカウント 001 と一番似ていたのはアカウント 007 であった。両アカウントの違いは 1. テーマが違ったキーワードを一回検索した、2. アカウント 007 がいいね・リツイート操作を抜いたの 2 点であった。またアカウント 004 と 007 の違いもいいね・リツイート操作である。すなわち、異なるキーワードを検索した場合、いいね・リツイートを抜いたアカウントは標準アカウントと一番似ていることがわかった。つまり、情報の繭からの脱出を希望する場合には見知らぬ領域のコンテンツへのいいね・リツイート操作は必要であることが分かった。

表1:第二段階実験のピアソン相関係数

相関係数	001	004	007	008
001	-	0.4472	0.8014	0.4953
004	-	-	0.6124	0.4101
007	-	-	-	0.4721

5. まとめ

実験結果より情報の偏りを予防するフォローチャートを提案した。今後の課題としてこのフローを Web アプリ等を実装することが挙げられる。

参考文献

- [1] Cass R. Sunstein: 『Infotopia: How Many Minds Produce Knowledge』, Oxford University Press, USA, 2008.
- [2] パリサーイーライ, 井口 耕二: 閉じこもるインターネット—グーグル・パーソナライズ・民主主義, 早川書房, 2012.
- [3] 門松健治, 遠藤玉夫, 岡 昌吾, 北川裕之/編: 実験医学増刊, 羊土社, Vol.31, No.10, 2013.