

Twitter上での会話の始まりやすさによる推薦のためのフォロー状態に関する考察

坪田 啓司† 小林 亜樹‡

†工学院大学

1 はじめに

SNSにおいて、友人の発見を支援するような方式として、アカウント間のプロフィール文、投稿文、あるいは「友人」関係グラフを用いて、興味嗜好の類似性を抽出する方式 [1][2] などがある。本研究では、Twitterにおいて必要なパラメータを入手することで、会話を容易に始められそうなアカウントを推薦する手法について提案している [3]。本稿では、改良した手法の提案と有効性の検証および、フィルタリングの考察を行う。

2 提案手法

2.1 概要

従来、相互フォローアカウントへの不偏的な会話行動を観測期間に観測して、始まりやすさの推定を行った。本稿では、新規の会話行動の引き起こしが直近の新規フォロー開始状況にも依存することを利用して、推定式による候補アカウント集合を絞り込み、推定値の高い順に会話の始まりやすいアカウントとして推薦する手法を提案する。既存の研究である推定値の導き方を2.2節にて、候補アカウント集合の絞り込みについてを2.3節にて説明する。

2.2 評価関数

推薦すべきアカウントであるかを評価する指標として、

1. フォローバックする割合
2. 会話ツイート相手の幅広さ
3. 会話相手へのツイート数の偏りのなさ

の3つを用いる。1は、推薦候補アカウント α のフォロワーの中で相互フォローアカウントの占める割合である。 α のフレンド集合を $f(\alpha)$ 、フォロワー集合を $F(\alpha)$ とすると、相互フォローアカウント集合 $\mathcal{F}(\alpha) = f(\alpha) \cap F(\alpha)$ である。よって、フォローバック率 $M(\alpha)$ は、

$$M(\alpha) = \frac{|\mathcal{F}(\alpha)|}{|F(\alpha)|} \quad (1)$$

である。

2は、 $\mathcal{F}(\alpha)$ のうち先としてツイートを行ったアカウントの割合である。 $\mathcal{F}(\alpha)$ において、観測期間内に α が行ったメンションツイートのあて先アカウント集合 $T(\alpha)$ を用いて、 $\mathcal{F}(\alpha)$ 中のツイートあて先となる割合 $R(\alpha)$ は、

$$R(\alpha) = \frac{|T(\alpha)|}{|\mathcal{F}(\alpha)|} \quad (2)$$

である。

3は、 $\mathcal{F}(\alpha)$ 内において、偏りなくメンションツイートを発しているかの指標である。 α が観測期間の終了時点における $\exists A_i \in \mathcal{F}(\alpha)$ への観測期間におけるメンションツイート数を $m(\alpha, A_i)$ とする。すると、この期間における α による $\mathcal{F}(\alpha)$ 全体へのメンションツイート数の合計 $m_{\mathcal{F}}(\alpha)$ は、

$$m_{\mathcal{F}}(\alpha) = \sum_{A_i \in \mathcal{F}(\alpha)} m(\alpha, A_i) \quad (3)$$

である。 $\forall A_i, p_i = \frac{m(\alpha, A_i)}{m_{\mathcal{F}}(\alpha)}$ として、 A_i へのメンションツイート数の $\mathcal{F}(\alpha)$ 全体に対する割合を定義することができ、これを用いて、 α によるツイートのあて先エントロピー $E(\alpha)$ を

$$E(\alpha) = - \sum_{i=1}^{|\mathcal{F}(\alpha)|} P_i \log P_i \quad (4)$$

と定義できる。

式(1), (2), (3)の全てが高い数値のアカウントを推薦するため、評価関数 $S(\alpha)$ を次のように定める。

$$S(\alpha) = M(\alpha)R(\alpha)E(\alpha) \quad (5)$$

2.3 推薦候補対象アカウントの絞り込み

2.2節において、 $M(\alpha)$ は観測期間内に新規に相互フォロー状態になったアカウントであるか否かは考慮していなかった。観測期間内で、新規に相互フォローを行う速度である相互フォロー速度がある程度ない場合、信頼性が低く、また、新規の人との交流を増やそうとしないと考えられる。

そこで、相互フォロー速度が一定値以下の場合に推薦候補集合から除外する。具体的には、観測期間前後での相互フォロー集合の差分、すなわち、新規相互フォローアカウント数 ΔI と、観測期間 d を用いて、

$$\Delta F_d = \frac{\Delta I}{d} \quad (6)$$

より、相互フォロー速度 ΔF_d を求め、 ΔF_d が一定値以下のアカウントを推薦候補集合から取り除く。取り除かれなかったアカウント集合を評価値 $S(\alpha)$ の降順に推薦アカウントとしていく。

3 評価実験

「バズドラ」, 「艦これ」, 「甲子園」の3つのいずれかのキーワードを発言していた計330アカウントを評価対象とした。これらのアカウントから観測期間において削除された、または鍵アカウントとなったもの、および新規相互フォローアカウント数が0のアカウントを除いた計256アカウントを推薦候補集合とした。これらの推薦候補集合に対して2013/8/21~9/29までの観測期間850時間において、相互フォローアカウント数が一定値以上のアカウントを取り除き、相互フォロー速度 ΔF_d が一定値以下のアカウントを取り除いた上で評価値 $S(\alpha)$ の高いアカウントを会話が始まりやすいアカウントとした。これにより継続的に会話を行うアカウントを推薦する。

観測期間に続く2013年9/29~10/22までの検証期間380時間において、新規に相互フォロー状態となったアカウントに対し、実際に会話行動をどの程度行っているかの割合との相関を見ることで有効性を検証する。また、異なる ΔF_d 閾値以下のアカウントを取り除いた推薦候補集合毎に両者の相関を計算した。

観測期間の終了時点における相互フォローアカウント数 O を横軸に、推定評価値 $S(\alpha)$ を縦軸とした散布図を図1に示す。図1における黄+緑点が推薦候補集合を示しており、緑点は $O \geq 700$ のアカウントを示している。図2は、 O を横軸に、検証期間において新規に相互フォロー状態となったアカウントに対してツイートする割合 μ_{γ} と推定評価値 $S(\alpha)$ との差を縦軸とした散布図である。図2における青+赤+黒+

A Study on the follow state for recommendation by the ease of the beginning of the conversation on Twitter

†Keishi Tsubota †Aki Kobayashi

‡Kogakuin University

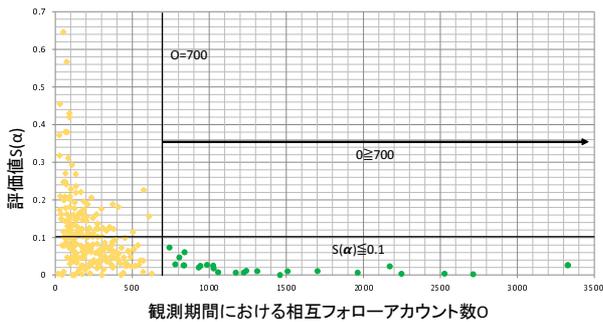


図 1: 相互フォローアカウント数と $S(\alpha)$ の関係

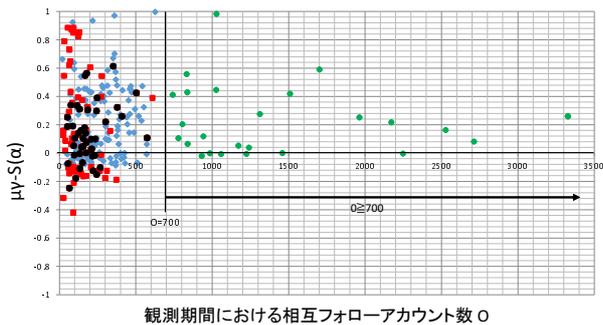


図 2: 相互フォローアカウント数と $\mu_\gamma - S(\alpha)$ の関係

緑点が推薦候補集合を示しており、緑点は $O \geq 700$ のアカウントを示している。

本稿は、継続的に会話を行うアカウントを推薦することが目的としているが、図 1,2 における緑点は妥当なアカウントではない。これらは、評価値 $S(\alpha)$ が $0 \leq S(\alpha) \leq 0.1$ に偏在し、 $\mu_\gamma - S(\alpha)$ の値が高いことから、検証期間において新規に相互フォロー状態となったアカウントに対して会話行動を行った率が高いことがわかる。一見、新規に相互フォロー状態となったアカウントに対して会話を行っている率が高いことから、会話を頻繁に行っているように見えるが、フォローされた際にあいさつ行動などを行う程度で継続的に会話を行っているわけではない。なぜなら、本研究の検証方法が新規に相互フォロー状態となったアカウントに対して会話行動を行ったかのみを測定しており、会話回数を考慮しないためである。したがって、見かけ上会話しているように見えるアカウントである $O \geq 700$ のアカウントは推薦候補集合として推薦すべきではない。 $O \geq 700$ のアカウントが $0 \leq S(\alpha) \leq 0.1$ に偏在していることから、 $0 \leq S(\alpha) \leq 0.1$ のアカウントを除外することで見かけ上会話しているように見えるアカウントを取り除くことができる。

次に、256 個の全推薦候補集合および、 $0 \leq S(\alpha) \leq 0.1$ のアカウントを取り除いた 101 個の推薦候補集合の 2 パターンにおいて、 ΔF_d が一定値以下のアカウントを取り除き、 $S\alpha$ と μ_γ との相関を計算する。

256 個の全推薦候補集合における推定評価値 $S(\alpha)$ を横軸に、検証期間において新規に相互フォロー状態となったアカウントに対してツイートする割合 μ_γ を縦軸とした散布図を図 3 に示す。異なる ΔF_d 閾値以下のアカウントを取り除き、相関を計算したところ $\Delta F_d = 0.07$ のときが最も相関係数が高くなった。

緑+赤+紫点が全評価対象アカウントで、相関係数は 0.264 である。紫点は $\Delta F_d = 0.07$ 以下のアカウントを取り除いた 21 個の最終推薦候補集合で、紫点における相関係数は 0.635 である。緑点は $O \geq 700$ のアカウントを示しているが、最終推薦候補集合である紫点にも含まれている。したがって、本フィルタリングによることで評価値による実際の会話行動の始まりやすさの推定精度は高まったと言えるが、一部の推薦すべきでないアカウントも推薦する結果となっている。

$0 \leq S(\alpha) \leq 0.1$ のアカウントを除外した 101 個の推薦候補集合について考える。 $O \geq 700$ のアカウントを緑点に、図 3 とは異なる ΔF_d の閾値以下のアカウントを赤点とし、最終推薦候補集合を黒点とした散布図を図 4 に示す。図 4 にお

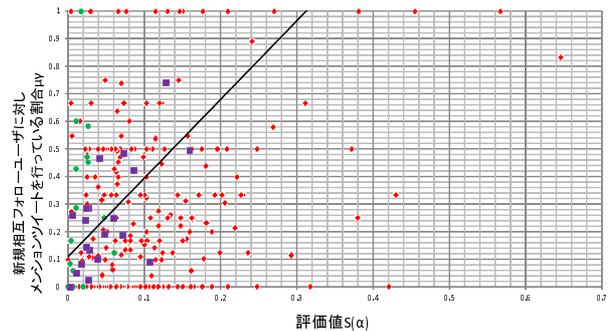


図 3: 256 アカウントにおける ΔF_d によるフィルタリング効果

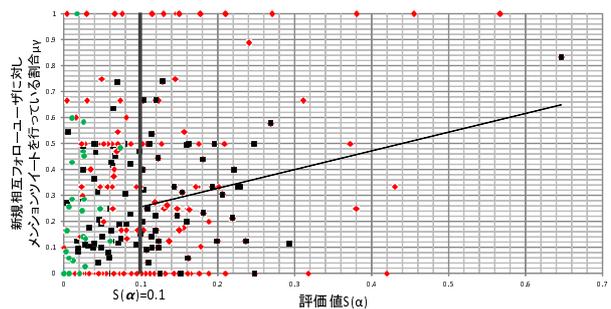


図 4: $S(\alpha) \geq 0.1$ における 101 アカウントにおける ΔF_d によるフィルタリング効果

いて、 $S(\alpha) = 0.1$ で推薦候補集合を区切り、 $S(\alpha) \geq 0.1$ における推薦候補集合における赤+黒点が 101 個の推薦対象アカウントで、相関係数は 0.298 である。 $S(\alpha) \geq 0.1$ における黒点は、 $S(\alpha) \geq 0.1$ のアカウントにおいて異なる ΔF_d 以下のアカウントを取り除いたときに相関係数が最大となった $\Delta F_d = 0.013$ 以下のアカウントを取り除いた 43 個の最終推薦候補集合で、相関係数は 0.31 である。

したがって、本フィルタリングによることで継続的に会話を行うアカウントを推薦することがある程度できた。しかし、 $O < 700$ かつ $0 \leq S(\alpha) \leq 0.1$ のアカウントを取り除いてしまった。これらのアカウントは図 2 における青点であり、 μ_γ が高いアカウントが含まれている。それらに含まれる推薦すべきアカウントも推薦することが今後の課題である。

4 おわりに

Twitter において、アカウントの行動を観測し、必要なパラメーターから会話の始まりやすさを推定し、相互フォロー速度が一定値以下のアカウントを取り除くことで会話が始まりやすいアカウントを推薦する手法を提案し、本手法の有効性を示した。フィルタリングにより省かれてしまった推薦すべきアカウントを推薦することが今後の課題である。

参考文献

- [1] 桑原 雄, 稲垣 陽一, 草野 奉章, 中島 伸介, 張 建偉: マイクロブログを対象としたユーザー特性分析に基づく類似ユーザーの発見および推薦方式, 情処研報, 2009-DBS-149(18), pp. 1-3(2009).
- [2] 古川 忠延, 松澤 智史, 松尾 豊, 内山 幸樹, 武田 正之: Weblog におけるユーザーのつながりと閲覧行動の分析, 電子情報通信学会論文誌. B, 通信 J88-B(7), 1258-1266, (2005).
- [3] 田村政人, 小林亜樹: Twitter における会話しやすいアカウントの推薦手法, 全国大会講演論文集 2013(1), 605-607(2013).