

Twitter におけるユーザの類似性に基づいた 個人化されたアイテムのランキングに関する検討

冨永 一成[†] 牛尼剛聡[‡]

[†]九州大学大学院芸術工学府 〒815-8540 福岡県福岡市南区塩原 4-9-1

[‡]九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]k.tominaga.238@s.kyushu-u.ac.jp, [‡]ushiana@design.kyushu-u.ac.jp

あらまし Web 利用の飛躍的な普及に伴って、我々は日常的に Web 上で映画、書籍、商品など様々なアイテムを検索するようになった。検索されたアイテム集合は、ユーザの評価等の客観的な基準でランキングされることが多いが、ユーザの主観的な情報は考慮されていない。本論文では、Twitter で店舗についてのツイートを投稿したユーザに注目し、ユーザ間の類似性にもとづいてアイテムの利用者の特徴を抽出する。そして、抽出された利用者の情報に基づいて、検索アイテムのランキングを作成するシステムを開発する。

キーワード ランキング, Twitter, ユーザプロフィール, 個人化

1. はじめに

Web 利用の飛躍的な普及に伴って、我々は日常的に Web 上で店舗、映画、書籍等の様々なアイテムを検索するようになった。検索されたアイテムが複数存在する場合、それらは何らかの基準でランキングされることが多い。多くのシステムでは主に、そのアイテムに対するユーザ評価等の客観的な基準を用いてランキングが行われていることが多い。しかし、そのようなシステムの多くは、システムを利用しているユーザの特性を考慮していない。例えば、店舗を選ぶ際、店舗に対する評価の他に、どのようなユーザがその店舗を利用しているのか、といった情報を加味することでより満足の行く検索を行うことができるのではないかと考えられる。また、ランキング形式でアイテムを提示する際、ユーザの興味がある可能性が高いアイテムでも、下位にランキングされると、見落とされたり、価値が低いと判断されてしまう可能性が高い。そこで本研究では、未知のアイテム集合に関して、ユーザの興味を反映したランキングを自動的に作成することを目的とする。

ユーザの興味に基づいたアイテム集合のランキングを自動的に生成するには、ユーザとそれぞれのアイテムとの適合度を推定する必要がある。本研究では、ランキング対象とするそれぞれのアイテムの特徴を抽出するために Twitter[1]を利用する。具体的には、Twitter 上でそれぞれのアイテムに関する記述を含むツイートを発信しているユーザを抽出し、そのユーザのツイート集合によってア

アイテムを特徴づける。あるアイテムに関するユーザ(そのアイテムについて述べているユーザ)のツイート集合と、システムを利用するユーザのツイート集合に基づいて、ユーザ間の類似度を推定し、その類似度に基づいて、アイテムのランキングを構成する。

従来の協調フィルタリングを始めとする推薦手法では、システムを利用するユーザのアイテムに対する評価等を必要とした。しかし、本研究で開発するシステムは、ユーザの特徴を Twitter で発言したツイート集合から取得するため、ユーザに新たに情報を入力してもらうことなく、未知のコンテンツの推薦を可能とする。また、「アイテムについてのツイートを発信しているユーザがどれだけ自分に似ているか」という新しい指標に基づいた推薦を行うことができる。

Twitter において、ユーザは様々な内容のツイートをしているが、ツイートはそのユーザの特徴を反映していると考えられる。そこで、本研究では、Twitter からユーザが発信したツイート集合を取得し、それらを特徴量としてユーザの類似度を判定する。Twitter は 140 字以内のテキストを投稿するという性質上、一つひとつのツイートは短くなりやすく、通常文書検索に用いられるキーワードに基づいた手法では価値のある結果を得にくい。そこで本研究では、書籍の著者推定に利用されている文字 n-gram 法が、ユーザの類似度として有効ではないかという仮説を立てた。その結果、文字 n-gram 法を用いることで、文の長さ、内容に関係なく、そのユーザ

独特の言い回しや特徴を抽出できると考えられるからである。

本論文では、この仮説に基づいて、4章でユーザの類似度を測るのに適切な特徴量を得るための予備実験を行い、5章で予備実験の結果を基にアイテムのランキングシステムの概要を示す。さらに6章で、映画をアイテムの例として、被験者の興味に基づいてアイテムをランキングしてもらった、システムが作成したランキングと、週間ランキングの2つを比較し、システムのランキングがユーザの興味を反映しているかを確かめる実験を行った。

2. 関連研究

著者推定の研究は古くから行われている。松浦ら[2]は、文字 n-gram の出現分布を手がかりに文章の著者を推定する手法を提案している。文章の構造や各単語の属性を考慮していないが、特定カテゴリに属する語の使用が特徴的である場合、その特徴が特定文字列の出現状況に全く反映されないとは考えにくい。よって、各文字 n-gram が文章間の疎密度に与えた影響の大きさに基づいて文体を特徴づけている文字列をより高度なレベルで把握することも可能であるとしている。本研究では松浦らが提案する関数 $dissim$ を、文字 n-gram を用いて Twitter ユーザ間の類似度測定のために利用する。

井上ら[3]は、品詞 n-gram を用いた著者推定手法を提案している。単語そのものではなく品詞を用いているため、異なる話題の文章群に対しても高精度な著者推定を期待することができるとしている。さらに、実験において既存の著者推定手法との推定精度の比較を行なっている。本研究では文字 n-gram との比較対象として品詞 n-gram を利用する。

中島ら[4]は、品詞と助詞の出現パターンを特徴として用いることで、記事の書き方が類似したブログ著者を推定し、類似著者のコミュニティとして抽出する手法を提案している。頻出するキーワードではなく、品詞と助詞の出現パターンを類似評価に用いることで、内容に偏りのないブログサイト同士の関係を発見することができるとしている。本研究はブログではなく、より類似度判定が難しいと考えられる Twitter のツイートを対象としている点が異なる。

3. アプローチ

本研究では、任意のアイテム集合に対して、ユーザの嗜好を反映したランキングを自動的に生成することを目的とする。

本手法では、Twitter で情報発信を行っているユーザを対象と

する。まず、ユーザ自身の特徴を取得するために、対象とするユーザが投稿したツイート集合を取得する。次に、ユーザが指定したアイテム集合それぞれに対して、Twitter 上でそのアイテムに関するツイートを投稿したユーザを抽出する。さらに、抽出したそれぞれのユーザについてツイートを取得する。それを基に、ユーザとの類似度を計算する。類似度に基づいて、アイテムのランキングを行い、結果をユーザに表示する。

4. ユーザの類似度の推定

4.1 n-gram を利用したユーザ類似度推定手法

本研究においては、Twitter 上での投稿に基づいてユーザの類似度に基づいて、アイテムのランキングを行う。

基本的には、ユーザが投稿したツイートに基づいてユーザの特徴を TF-IDF 法を用いたツイートに含まれる単語の重みによって表現することが考えられる。Twitter においてユーザは、「ラーメンなう」に代表されるように、日常で起こった出来事を短い文章で発信することが多い。例えば映画を観た際、「今日〇〇を観た！面白かった！」といったツイートを投稿する場合がある。しかし、bot や宣伝用のアカウント等、一部のユーザを除けば、ユーザが特定のキーワードを含むツイートを発信することは、1, 2 回程度である。そのため、ユーザの類似性を判定する使用量として、TF-IDF 等の単語の頻度に注目した特徴量を利用することは適切でないと考えられる。

そこで本研究では、文字 n-gram の出現頻度を利用することを考える。文字 n-gram 法を用いれば、文の長さ、内容に関係なく、そのユーザ独特の言い回しや特徴を抽出できると考えられ、話題に依存しない類似性の指標として適切であると考えられるからである。抽出した特徴から、ユーザの類似度を測ることができれば、ユーザの特性を考慮した推薦をすることが可能になる。

文字 n-gram 法を用いたユーザの類似度を測る関数として、松浦ら[2]が提案した $dissim$ を応用する。いま、 x を長さ n の文字列 (n-gram) とし、文書集合 P および文書集合 Q における x の出現確率分布をそれぞれ $P(x)$, $Q(x)$ とすると、tweet 集合 P および Q 双方に出現する n-gram、すなわち共通 n-gram 集合 C は以下で表される。

$$C = \{x \mid P(x)Q(x) \neq 0\}$$

ここで、集合 C の要素数を $card(C)$ とすると、 $dissim$ は以下のように計算できる。

$$\text{dissim}(P, Q) = \frac{1}{\text{card}(C)} \sum_{x \in C} |\log(P(x)/Q(x))|$$

いま、ユーザ u が発したツイート集合を $TW(u)$ と表記するとき、ユーザ u_1 及び u_2 間の距離 $\text{dissim}U(u_1, u_2)$ を以下のように定義する。

$$\text{dissim}U(u_1, u_2) = \text{dissim}(TW(u_1), TW(u_2))$$

比較する tweet 集合双方に出現する文字列のみに着目することで、文章間の疎遠度が内容の違いではなく、文章の癖の差によって決定されると考えられる。

4.2 ユーザ類似度に関する実験

ユーザの類似度に有効な特徴量を得るために 3 つの特徴量に関する比較実験を行った。実験に使用した特徴量は文字 2-gram, TF/IDF, 品詞 2-gram である。

4.2.1 比較に使用する特徴量

ここで、TF-IDF と品詞 n-gram について説明する。

まず TF-IDF について説明する。単語 i の文書 j に於ける重み $\text{TFIDF}(i, j)$ は以下の式で表される。

$$\text{TF-IDF}(i, j) = \text{TF}(i, j) \times \log\left(\frac{N}{\text{DF}(i)}\right)$$

ここで、 $\text{TF}(i, j)$ は、キーワード i がドキュメント(ツイート) j 中の出現頻度を表す。 N は全てのドキュメントの数を表す。 $\text{DF}(i)$ はキーワード i が含まれるドキュメント数を表す。なお、ここでは、各ユーザにつき 1000 件のツイートを 1 つのドキュメント集合として扱う。

この計算式に基づいて各ユーザのツイート集合における単語の重みを決定する。

ユーザ間の類似度の計算にはコサイン相関値を利用する。いま、ユーザ u_1 のツイート集合における単語の重みベクトルを a 、ユーザ u_2 の単語の重みベクトルを b とする。このとき、ユーザ u_1 と u_2 の類似度は、それぞれのユーザの特徴ベクトルのコサイン相関値として次の式で表される。

$$\text{sim}(u_1, u_2) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^n a_i \cdot b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}}$$

品詞 n-gram とは、品詞の組み合わせを特徴量とするものである。例えば「今日は走った」という文章ならば、「今日」「は」「走った」と分解でき、さらに「名詞」「助詞」「動詞」となる。2-gram であれば、「名詞助詞」「助詞動詞」が特徴量となる。ユーザとの類似度を測る関数は 2-gram と同様に dissim を用いてユーザ間の類似度を計

算する。

4.2.2 実験の内容

実験内容と結果を以下に示す。

- TwitterAPI を用いて、Twitter からユーザ 100 人をランダムに抽出
- 抽出したユーザが投稿したツイートを 1000 件程度抽出
- ツイートを 2-gram, TF/IDF, 品詞 2-gram で解析し、被験者との類似度を計算
- それぞれのランキング手法に於いて、ユーザをランキングし、1 位, 10 位, 20 位, 30 位, 40 位, 50 位, 60 位, 70 位, 80 位, 90 位の 10 件のユーザを取得
- 取得したユーザをランダムでユーザに提示し、それぞれ 5 段階(5:似ている, 4:まあ似ている, 3:どちらでもない, 2:あまり似ていない, 1:似ていない)で評価

4.2.3 実験結果

実験では、5人の被験者に評価をしてもらった。結果を示

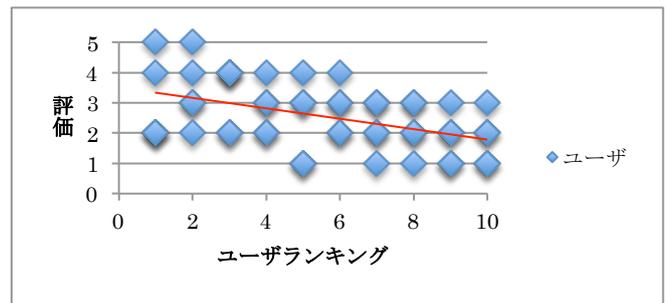


図1 文字 2-gram を利用した場合の相関係数のグラフ

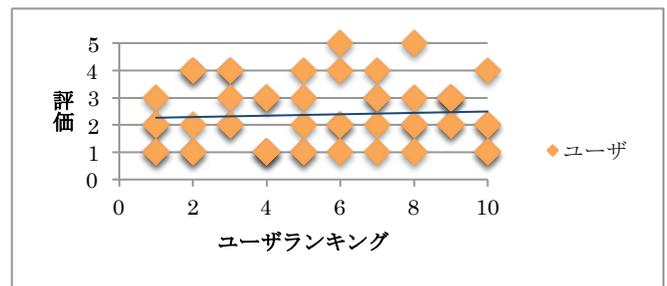


図2 TF-IDF を利用した場合の相関係数のグラフ

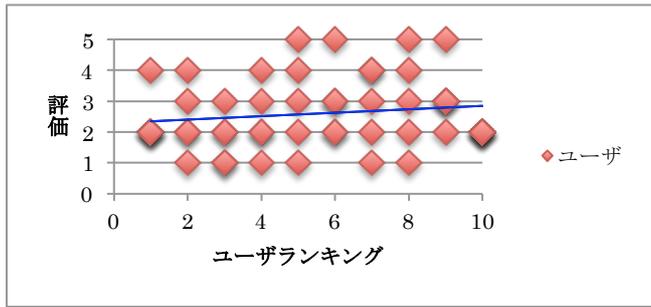


図3 品詞 2-gram を利用した場合の相関係数のグラフ

表1 ユーザ類似度に関する相関係数

	文字 2-gram	TF/IDF	品詞 2-gram
被験者 1	-0.79	-0.38	0.20
被験者 2	-0.60	-0.63	-0.12
被験者 3	-0.67	0.11	0.076
被験者 4	-0.16	0.63	0.044
被験者 5	0.07	-0.22	-0.029
総合評価	-0.46	0.06	0.14

したグラフの横軸は関数 $dissimI$, コサイン相関値ともに相対的な値であるため, システムが自動的に作成したユーザーランキングを類似度が高い順に 1~10 位でランク付けしたものを使用した. 表は各被験者の評価とシステムが作成したランキングとの相関係数をまとめたものである.

文字 2-gram と品詞 2-gram を利用してユーザーの類似度を測る関数 $dissimU$ は 0 に近いほどユーザーが類似していることを表す. よって被験者の評価との間に負の相関があればシステムの精度が高いことを示す. TF/IDF を利用して類似度を測る関数コサイン相関値は, 1 に近いほどユーザーが類似していることを示す. よって $dissimU$ とは逆に正の相関があればシステムの精度が高いことを示す.

次に結果に関する考察を述べる. 文字 2-gram に関して, 3 人の被験者に -0.6 以上の負の相関を見ることができた. TF-IDF に関しては, 一人の被験者に 0.6 以上の正の相関を見ることができたが, -0.6 以上の負の相関も見られた. 品詞 2-gram に関しては負の相関は全く見ることができなかった. 被験者全体として見ても文字 2-gram に関して一定の負の相関を見ることができたが TF-IDF と品詞 2-gram には相関は見られなかった. 以上から, 文字 2-gram がユーザーの類似度を測る特徴量として最も適切であること

が分かった.

5. ツイッターに於けるユーザーの類似度に基づくアイテムのランキング手法

Twitter に於けるユーザーの類似度に基づいて, 対象とするユーザーにとってのランキングを行う手法を提案する. 提案手法の大まかな流れを以下に示す(図 4).

1. システムにランキングの対象と成るアイテム集合を与える.
2. それぞれのアイテムに関して, Twitter 検索 API を用いてそのアイテムに関するツイートを投稿したユーザー集合を取得する.
3. 2 で取得した twitter ユーザそれぞれに対して, Twitter API を用いてそのユーザーが投稿したツイートを 1000 件取得する.
4. システムを利用するユーザーと, 2 で取得したユーザー集合との類似度に基づいてアイテムの適合度を求める.
5. 全てのアイテムの適合度を計算し, それに基づいてランク付けする.

ここで, ユーザー u に対するアイテム $item$ の適合度 $score(u, item)$ は, 以下の式で求める.

$$score(u, item) = \sum_{x \in users(item)} dissimU(u, x)$$

ここで, $users(item)$ は $item$ についてのツイートを投稿した Twitter ユーザーである. $dissimU$ の値が小さいほどユーザーの類似度が高いと考えられるため, $score$ の値が小さい $item$ ほど上位にランク付けする.

6. 実験

提案手法の有効性を示すため, システムのプロトタイプを実装し, 実験を行った. ランキングするアイテムの集合としては映画を用いた. 具体的には, 映画.com の 2013 年 1 月第 2 週の週間映画ランキング上位 10 件の映画を対象として, 各映画のタイトルを含むツイートを発信しているユーザー 27 人, 合計 270 人のツイートをそれぞれ約 1000 件ずつ取得した. 尚, Twitter API の制限のため, 一度に取得できるユーザーのツイート数は限られている. そこで, 本実験ではリアルタイムにデータを取得するのではなく, 予め用意したデータ・セットを用いた.

被験者5人に対して、実験対象とする映画10件に興味がある順に映画をランキングしてもらい正解データとする。そして、提案手法に基づいて得られたランキングを求め、精度を比較した。また、ベースラインとして、観客動員数に基づく週間ランキングを用い、提案手法の性能を比較した。

表2に、あるユーザによって行われたユーザ評価のランキングと提案手法によるランキングの結果の例を示す。

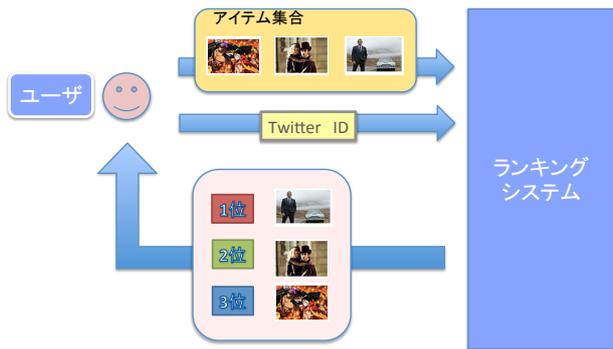


図4 システム概要

表2 被験者1のランキング結果

	ユーザ	提案手法	週間 ランキング
レ・ミゼラブル	1	1	2
スカイフォール	2	3	3
大奥	3	6	8
フランケンウィニー	4	2	10
今日、恋をはじめます	5	5	6
青の祓魔師	6	7	9
ホビット	7	4	5
ONE PIECE FILM Z	8	10	1
妖怪人間ベム	9	8	4
仮面ライダー	10	9	7

表3 MAEの結果

	提案手法	週間 ランキング
被験者1	1.4	3.4
被験者2	1.8	3.8
被験者3	2.2	2.8
被験者4	2.8	2.8
被験者5	3.0	3.2
被験者全体	2.24	3.2

精度の評価には、絶対平均誤差(mean absolute error; MAE)を用いた。MAEとは、特定のアイテムに対するランキングを、システムが正確に予測する能力を評価する際に最もよく用いられる指標である。MAEは以下の式で表される。

$$MAE = \frac{\sum_{u \in U} \sum_{i \in testset_u} |rec(u, i) - r_{u,i}|}{\sum_{u \in U} |testset_u|}$$

MAEは評価を行ったすべてのユーザ $u \in U$ とテストデータセット ($testset_u$)内のすべてのアイテムに対して、計算された推薦スコア ($rec(u, i)$)と実際の評価値 ($r_{u,i}$)の間の偏差の平均を計算する。本実験では、ユーザが作成したランキングを基準としてMAEを算出した。

表3に提案手法および週間ランキングのMAEの結果を示す。

提案手法により自動的に作成したランキングのMAEは、週間ランキングよりも、ユーザの興味との誤差が少ないという結果を得た。特に、被験者の内2名においては、システムのMAEが週間ランキングとのMAEの差が顕著であった。ここで、さらにシステムと週間ランキングのMAEの間に有意差があるかを調べるためにt検定を行った。結果を以下に示す。

P	0.017
---	-------

$p = 0.017 < 0.05$ であるので、5%水準で有意差があると言える。

以上より、本提案手法によっても求められるランキングは、週間ランキングよりもユーザの興味に近いランキングを提示することができた。

しかし、システムがランキングしたアイテムを個別に見ると、下位

にランキングされやすいアイテムが見られた。これは、Twitter からユーザを取得する際、bot や宣伝用のアカウント等、被験者との類似度が低いと思われるユーザが多数含まれていたからではないかと考えられる。今後は、それらのユーザをできるだけ除去し、よりランキングの精度を上げていくことが課題である。具体的には、システムが算出したユーザの類似度 $dissimU$ が閾値よりも大きい場合、システムに有効でないユーザと判断する等が考えられる。

6. まとめ

本論文では、Twitterにおけるユーザの類似度に基づいて、ユーザの興味を反映したアイテムのランキングを作成する手法を提案した。

文字 n-gram 法の有用性を確かめる予備実験では TF-IDF、品詞 n-gram との比較を行い、文字 n-gram がユーザとの類似度を測る特徴量として最も有効であることを確かめた。つぎに、Twitter ユーザの類似度に基づいて、ユーザに適合したアイテムのランキングを行う手法を提案した。

提案手法のプロトタイプシステムを作成し、映画をアイテム例とした実験を行った。実験では、被験者の興味に基づいて予めアイテムをランキングしてもらい、システムが自動的に作成したランキングとの誤差を測ることで精度を測った。比較対象として週間ランキングを用い、システムが作成したランキングが、よりユーザの興味を反映したランキングであることを確かめた。

今後は、システムの精度を改善するとともに、地域の飲食店等をランキングするアイテムとして扱うことを課題とする。初めて訪れた土地で、自分と嗜好が似ている人が話題にしている店舗ランキングを提示することで、ユーザの選択肢を広げることが期待でき、本システムを使用する状況として、最も有効であると考えられるからである。

参考文献

- [1] <http://twitter.com>
- [2] 松浦司, 金田康正: n-gram 分布を用いた近代日本語小説の著者推定, 軽量国語学, vol.22, no6, pp.225-238
- [3] 井上雅翔, 山名早人: 品詞 n-gram を用いた著者推定手法一話題に対する頑健性の評価, DBSJ Journal, vol.10, no3, pp7-12

[4] 中島 泰, 山名 早人: 品詞と助詞の出現パターンを用いた類似著者の推定とコミュニティ抽出, DEIM Forum 2011