

ユーザ影響度分析法によるレビューデータ特性分析 Characterizing Review Date by user influence Analysis Method

鈴木 優人[†]
Yuto Suzuki

斉藤 和巳[†]
Kazumi Saito

1. はじめに

近年、商品やサービスについてのレビューを投稿できるレビューサイトを利用するユーザが増えており盛んに活用され、レビューサイト分析や研究が多く展開されている [Salganik 06]。これらレビューサイトにおけるユーザのレビューは、他のユーザの購買活動に強く影響を与えるものであり、レビューサイトは重要なメディアといえる。その中でも、他のユーザに影響を与えるユーザ抽出技術は重要な研究課題である。これまでの我々の研究では、6つのレビューサイトを用いてユーザのスコアを計算し、影響力のあるユーザを分析した。分析結果より、ユーザごとに影響度が異なるとしたユーザモデルと、レビュー順序ごとに影響度が異なるとしたオーダモデルはレビューサイトごとに異なる傾向を示すことを確認した [Suzuki 16]。

そこで本研究では、これまで調査してきた影響力のあるユーザは実際にはどのような特徴があるのか3つの指標を利用して評価する。投票者モデル [Even-Dar 07] に基づき、6つのレビューサイトを用いて検証する。

2. 提案指標

ユーザ集合を $U = \{u_1, \dots, u_N\}$ 、アイテム集合を $I = \{i_1, \dots, i_M\}$ とし、ユーザ u のレビューしたアイテム集合を $I(u) \subset I$ 、アイテム i にレビューしたユーザ集合を $U(i) \subset U$ とする。また、ユーザ u のアイテム i に対する評点を $s(u, i)$ 、レビューした時刻を $t(u, i)$ とし、アイテムに対する評点は $S = \{s_1, \dots, s_K\}$ のいずれかの値とする。

本稿では、レビューサイトにおいて影響力のあるユーザの特性を分析するため以下の3つの指標を用いる。1つ目の評価指標として、ユーザ u の総レビュー数を用いる。ユーザ u の評価アイテムの総数は $|I(u)|$ で定義される。

2つ目の指標として、ユーザのレビューの影響範囲を計る指標 $Succ(u)$ を用いる。 $Succ(u)$ は、ユーザ u がアイテム i をレビューした時刻 $t(u, i)$ より後にレビューしたユーザ数の平均値にて定義され、ユーザ u がアイテム i にレビューした時刻よりも後にレビューしたユーザの集合を $U(i, u)$ とすると、指標 $Succ(u)$ は以下の式で定義される。

$$Succ(u) = \frac{1}{|I(u)|} \sum_{i \in I(u)} |U(i, u)| \quad (1)$$

3つ目の指標として、ユーザの評点が後続のユーザの評点行動に影響を与えているかどうかを計る指標 $Rsim(u)$ を投票者モデルを用いて定義する。 $Rsim(u)$ は、ユーザ u の各アイテム i に対する評点 $s(u, i)$ と、

ユーザ u より後にアイテム i をレビューした他のユーザ v のうち、何人が同じ評点を付けたかどうか、つまりは $s(u, i)$ と $s(v, i)$ の一致度より定義される。 $U(i, u)$ の中で評点 s を与えたユーザ集合を $U(i, u, s) \subset U(i, u)$ とすると、ユーザ u の $Rsim(u)$ は以下の式で定義される。

$$Rsim(u) = \frac{1}{|I(u)|} \sum_{i \in I(u)} \frac{|U(i, u, s(u, i))|}{|U(i, u)|} \quad (2)$$

投票者モデルは、ユーザが近傍のユーザの意見に基づいて評点を変更・決定する確率過程のモデルであり、モデルに従えば、後続ユーザ群がユーザ u のレビューを参考にした、または影響を受けた場合、後続ユーザはユーザ u と同じ評点を付けると考えられる。よって、後続ユーザ群の多くが同じ評点を付けているほど、ユーザ u のレビューは他のユーザに対して影響を及ぼしたといえる。また、この指標により、極端な主観を持ち、度が過ぎる肯定や否定をするレビューを投稿するユーザは低い評価を受けるため、このようなユーザを弾く役割も持つ。

本研究では、ユーザのレビュー影響力を定量的に求めるために、各ユーザの重みをパラメータ $X = \{x_u, \dots\}$ として設定し、アイテム i にユーザ u が評点 s を与える確率をモデル化し、パラメータ集合 X を最尤推定で求める。すなわち、先述するように、アイテム i にユーザ u より前にレビューしたユーザ集合を $U(i, u)$ とし、 $U(i, u)$ の中で評点 s を与えたユーザ集合を $U(i, u, s) \subset U(i, u)$ とすると、アイテム i にユーザ u が評点 s を与える確率は次のようにモデル化できる。

$$P(i, u, s; X) = \frac{\sum_{v \in U(i, u, s)} x_v}{\sum_{v \in U(i, u)} x_v} \quad (3)$$

そして、収集したレビューデータ D で、実際にアイテム i に対しユーザ u が評点 s を付けたことより、次式の尤度 $L(X)$ を最大化するパラメータ X を求める。

$$L(X) = \sum_{(i, u, s(u, i)) \in D} \log P(i, u, s(u, i); X) - \frac{r}{2} \sum_{u \in U} \log(x_u)^2 \quad (4)$$

ここで r は正則化係数を表し、特に過学習を防ぐために用いる。 $L(X)$ の最大化については、任意のユーザ u のパラメータを $x_u = 1$ として初期値 $X^{(0)}$ を設定し、EM アルゴリズムの第 n 反復目での尤度の改善が $L(X^{(n)}) < (1 + \epsilon)L(X^{(n-1)})$ となるまで学習させた。なお実験では、 $\epsilon = 0.0001$ に設定した。

3. 実験設定

実験には6つのレビューデータを用いる。アニコレは、アニメを題材としたランキング、口コミサイト、@cosme は、化粧品をアイテムとしたレビューサイトであり、ブックログは、ブックレビューサイト、食ベログは、全国飲食店の口コミ、ランキングサイトであり、価格.com は、あらゆる製品を比較検討できる買い物支援サイト、MovieLens は、映画の評価がまとめられている。表1にこれらレビューサイトのデータ概要のまとめを示す。

[†]静岡県立大学 経営情報学部

表 1: レビューサイト詳細表

	アニコレ	@cosme	ブックログ	食べログ	価格.com	MovieLens
アイテム数	1,790	46,398	178,432	443,573	75,519	10,677
ユーザ数	13,111	10,403	157,325	291,052	170,645	69,878
レビュー数	300,327	297,453	2,752,817	3,114,507	475,327	10,000,054

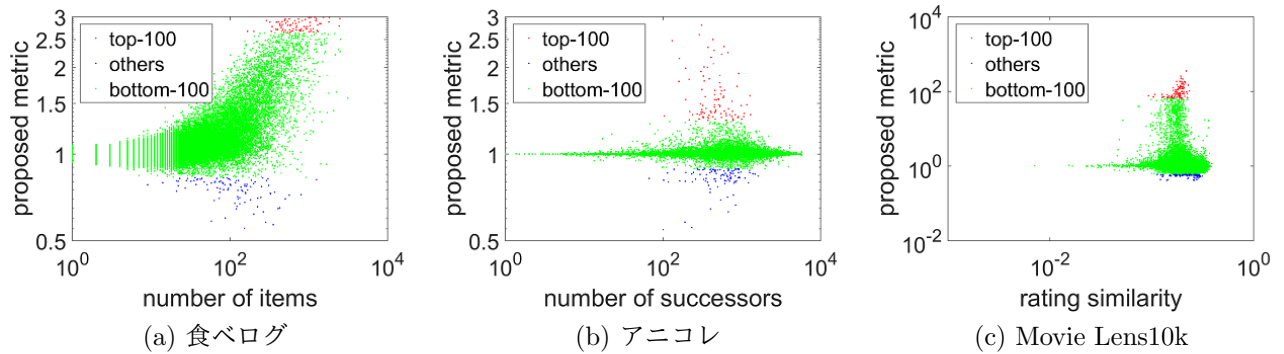


図 1: 各サイトレビュー数の相関図

4. 実験結果

図 1(a) は縦軸をユーザ重みパラメータ x_u , 横軸を評価アイテム数 $|I(u)|$ とし, ユーザをプロットしたものである. 図では, ユーザ重み上位 100 のユーザ, 下位 100 のユーザ, その他のユーザをそれぞれ, 赤点, 青点, 緑点で示している. 例としてここでは食べログの実験結果を示す. 見て取れるよう右上に広がっている. これによりユーザ影響度の高いユーザはレビュー数が多いことがわかる. 他のレビューサイトの結果をみても 1 つ目の評価指標のアイテム数では若干の相関が見られた.

図 1(b) は縦軸をパラメータ x_u , 横軸を $Succ(u)$ としている. ここではアニコレの実験結果を示す. 影響力のあるユーザはある程度早期のレビューをしていることがわかるがあまりにも早い時期にレビューしているユーザは総じて影響力が低いこともわかる. これは, 早すぎるレビューは主観的な意見であったり, 少数の一部ユーザの意見しか反映されないため, 客観性に欠け, ある程度の意見を踏まえたレビューが望まれていると考えられる. つまり, 先のレビュー者の良い, 悪い意見を踏まえた上での評点が望まれており, そのようなレビューが後続ユーザ群へ強い影響を与えられられる. なお, 他のレビューサイトの結果をみても似たような傾向が見られた.

図 1(c) は, 縦軸をパラメータ x_u , 横軸を $Rsim(u)$ としている. ここでは例として Movie Lens の実験結果を示す. Movie Lens は他のレビューサイトと比べても分かるが, パラメータ x_u の値が極端に他のレビューサイトと違っている. しかし, レビューの早さと一致度の観点からは, 他のレビューサイトの結果と似たような傾向を示した. また一致確率が高すぎる, $Rsim(u)$ が極端に高いユーザは, レビュー時期が遅い, レビュー数が少ない等の原因で発生しているものだと考えられる.

5. おわりに

本研究では, これまで調査してきた影響力のあるユーザは実際にはどのような特徴があるのか 3 つの指標を利用して 6 つのレビューサイトで評価した. 合計 18 個の実験結果より, 調査してきた影響力のあるユーザはレビューしているアイテム数も多く, 早すぎないがある程度早期のレビューをし, 後続ユーザのレビュー点数の一致確率がとても高いことが分かった. 今後は多様なレビューサイトを用いて, 活動レベルなどユーザー属性と影響度との関係分析なども進めたい.

謝辞 本研究は, 科学研究費補助金基盤研究(C)(No.17H01826)の助成を受けた.

参考文献

- [Salganik 06] M.J.Salganik, P.S.Dodds, and D.J.Watts, "Experimental Study of Inequality and Unpredictability in an Artistic Cultural Market", Science 10, pp.854-856, February 2006, (2006).
- [Even-Dar 07] Even-Dar.E. and Shapira.A, "A note on maximizing the spread of influence in social networks", In Proc. of WINE 2007, pp.281?286, 2007, (2007) .
- [Suzuki 16] 鈴木 優人, 齊藤 和巳, "投票者モデルに基づく複数レビューサイトでの影響度分析", 第 15 回情報科学技術フォーラム (FIT2016)