

投票者モデルに基づくユーザ影響度分析 Analysis of user influence based on voter model

永倉 卓弥[†]
Takumi nagakura

小林 えり[†]
Eri Kobayashi

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言える。しかし、多くの情報の中から参考になるレビューやユーザを発見することは容易ではない。例えば、多くのユーザから信頼を得ているユーザ（スペシャリスト）や、ブームや商品の状態変化（改良、改悪）等の外部要因をいち早くサイトに持ち込む情報に敏感なユーザ（Early Adopters）のレビューは他のユーザに強い影響力を持つと考えられ、そのようなユーザ抽出技術は重要な研究課題といえる。

そこで本研究では、投票者モデル [Even-Dar 07] に基づき、3つの指標を用いてユーザのスコアを計算し、レビューサイトにおいて影響力のあるユーザの抽出を行う。

2. 提案指標

レビューサイトにおいて影響力のあるユーザの抽出に関して、以下の3つの指標を用いて計算される指標 $P(u)$ を用いて抽出する。

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。

一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 $U(i, t(u, i))$ とすると、指標 $Succ(u)$ は以下の式で定義される。

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

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。ユーザ u の $Rsim(u)$ は以下の式で定義される。

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

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。

本稿では上記3つの指標を組合せて新たに以下の指標を提案する。

$$P(u) = \sum_{i \in I(u)} |v \in U(i, t(u, i)); s(v, i) = s(u, i)| \quad (3)$$

提案指標値 $P(u)$ の高いユーザは早期に、多くのアイテムをレビューし、他のユーザに対して影響力を持つレビューを投稿するようなユーザとなるため、重要ユーザの抽出が期待できる。

3. 実験設定

実験には@cosmeを用いる。@cosmeは、化粧品をアイテムとしたレビューサイトであり、2008年12月から2009年12月にかけてクロールして取得したものである。48,548アイテム、26,804ユーザ、297,453レビューを有する。ただし、ユーザは一回以上レビュー行動をとったユーザに限定する。レビューの得点は、1~7の整数値をとりうるため、 $S = \{1, \dots, 7\}$ となる。

4. 実験結果

図1に各指標間の相関を示し、図1(a)にて $I(u)$ と $Rsim(u)$ 、(b)にて $I(u)$ と $Succ(u)$ 、(c)にて $Succ(u)$ と $Rsim(u)$ での相関図を示す。例えば、図1(b)より、評価アイテム数 $I(u)$ が1つしかないユーザの $Succ(u)$ は高い評点もあれば低い評点もあり、これはたまたまレビューした時期が早かった、遅かったの違いによるものと分かる。

図1(a)、(b)を見ると、アイテム数の増加に伴いユーザの $Rsim(u)$ 及び、 $Succ(u)$ 評点は平均に近づいていくことが確認できる。一方、図1(c)より、 $Rsim(u)$ と $Succ(u)$ 間に相関関係は見られなかった。同じ投稿時期 ($Succ(u)$ 値) でも $Rsim(u)$ 値の高い、低いユーザの存在が確認でき、極端な主観にて評価を行うユーザがいることが分かった。

図2に各指標と提案指標 $P(u)$ による相関図を示す。図2に関して、(a)は $I(u)$ 、(b)は $Succ(u)$ 、(c)は

[†]静岡県立大学 経営情報学部

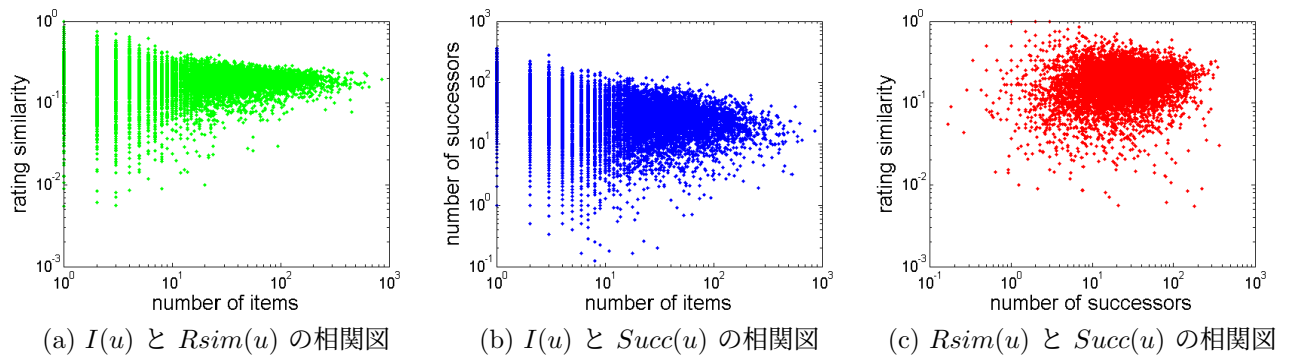


図 1: 各指標の相関図

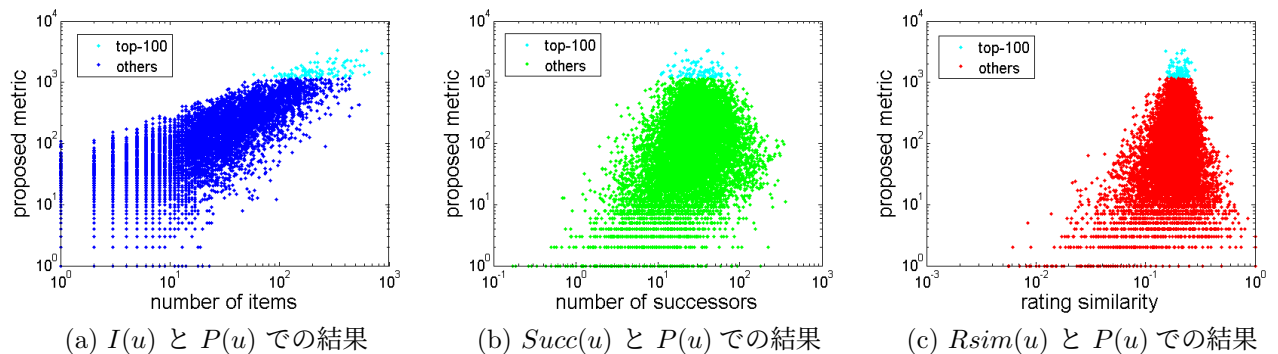


図 2: 各指標と提案指標の相関図

$Rsim(u)$ を横軸にとり、縦軸は 3 つとも $P(u)$ 値をとる。図中の点に関して、 $P(u)$ 値の高い上位 100 ユーザは 3 つとも共通して水色で示す。

図 2(a) より、評価アイテム数 $I(u)$ の増加に伴い、高い $P(u)$ 値を確認でき、評価レビュー数が提案指標に強い影響を与えていることが分かる。図 2(b) に関して、 $Succ(u)$ と $P(u)$ 間に相関関係は見られず、水色で示す上位 100 ユーザ群は $10 \sim 100$ ほどの $Succ(u)$ 値に分布している。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、先のレビューの悪い、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ると考えられる。図 2(c) に関して、図 2(b) 同様、 $Rsim(u)$ と $P(u)$ 間には相関関係は見られなかった。上位 100 ユーザ群と同じ $Rsim(u)$ 値であっても $P(u)$ 値の低いユーザが多く見られ、これはユーザの総レビュー数、レビュー投稿の時期の違いが原因である。

以上より、提案指標により、多くのアイテムにレビューし、かつ、ある程度の時期（ある程度レビュー数がたまった時期）にレビューをし、そのレビューが他のユーザに影響を与えるようなユーザの抽出が期待できる。このようなユーザはアイテムの良し悪しを踏まえた妥当な評点を下すことが期待でき、後続ユーザの良いアンカーとなる。よって、あるアイテムの有力な参考情報として、このようなユーザのレビュー情報の提示が役立つと考えられる。

5. おわりに

本稿では、他のユーザのレビュー評点行動に影響を与える重要ユーザ抽出を目的に、3 つの指標を基に計算される新たな指標を提案した。実データを用いて評価したところ、提案指標により、アイテムに対し妥当な評点を下すユーザの抽出が期待できることを確認した。今後は提案指標値の上位ユーザの分析、他のレビューデータでの検証を行う。また、 $Rsim(u)$ に関して、例えば、評点の変化点検出技術を取り入れるなど、時系列情報を取り入れた後続ユーザへの影響度を計る指標の提案を行う。

謝辞 本研究は、総務省 SCOPE(No.142306004)、及び、科学研究費補助金基盤研究 (C)(No.15K00429) 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 Artificial 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).