

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 Influence Analysis of Multiple Review Sites based on Voter Model

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ー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ーのレビューは、他のユーザー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ーを発見することは容易ではない。例えば、多くのユーザーから信頼を得ているユーザー (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ー (Early Adopters) のレビューは他のユーザーに強い影響力を持つと考えられ、そのようなユーザー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、6つのレビューサイトを用いてユーザーのスコアを計算し、レビューサイトにおいて影響力のあるユーザーの抽出を行う。

2. 提案指標

ユーザー集合を $U = \{u, v, \dots\}$, アイテム集合を $I = \{i, \dots\}$ とし、ユーザー 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ー集合を $U(i) \subset U$ とする。また、ユーザー u 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 $S = \{s, \dots\}$ の中のいずれかの値とする。

レビュー影響の算出には、以下の2つ方法を用いる。第一の方法では、ユーザー重みをパラメータ $X = \{x_u, \dots\}$ として設定し、アイテム i にユーザー 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、アイテム i にユーザー u より前にレビューしたユーザー集合を $U(i, u)$ とし、 $U(i, u)$ の中で評点 s を与えたユーザー集合を $U(i, u, s) \subset U(i, u)$ とすると、アイテム i にユーザー u が評点 s を与える確率は次のようにモデル化できる。

$$P(i, u, s; X) = \frac{\sum_{v \in U(i, u, s)} x_v}{\sum_{v \in U(i, u)} x_v} \quad (1)$$

そして、収集したレビューデータ D で、実際にアイテム i に対しユーザー 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。

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

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、任意のユーザー u のパラメータを $x_u = 1$ とし初期値 $X^{(0)}$ 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。以下、この方法はユーザー重み付け法 (user 法) とする。

第二の方法では、レビュー順序集合 $H = \{1, \dots, h, \dots\}$ に重みを与えたパラメータ集合 $Y = \{y_1, \dots, y_h, \dots\}$ として設定し、アイテム i に対し h 番目に評点 s を与える確率をモデル化し、パラメータ集合 Y を最尤推定で求める。すなわち、順序 h より前のレビュー順序の集合を $H(h) = \{1, \dots, h-1\}$ とし、 $H(h)$ の中で評点 s を与えたレビュー順序の集合を $H(h, s)$ とすると、アイテム i に対し h 番目に評点 s を与える確率は以下のようにモデル化できる。

$$P(i, h, s; Y) = \frac{\sum_{j \in H(h, s)} y_j}{\sum_{j \in H(h)} y_j} \quad (3)$$

また、ユーザー重み付け法と同様に、収集したレビューデータ D で、実際にアイテム i に対し h 番目に評点 $s = s(i, h)$ を付けたことより、次式の尤度 $L(Y)$ を最大化するパラメータ Y を求める。

$$L(Y) = \sum_{(i, h, s(i, h)) \in D} \log P(i, h, s(i, h); Y) - \frac{r}{2} \sum_{h \in H} \log(y_h)^2 \quad (4)$$

この方法はレビュー順序重み付け法 (order 法) とする。

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、口コミサイトであり、1,790 アイテム、13,111 ユーザー、300,327 レビューを有する。@cosme は、化粧品をアイテムとしたレビューサイトであり、46,398 アイテム、10,403 ユーザー、297,453 レビューを有する。ブックログは、ブックレビューサイトであり、178,432 アイテム、157,325 ユーザー、2,752,817 レビューを有する。食べログは、全国飲食店の口コミ、ランキングサイトであり、443,573 アイテム、291,052 ユーザー、3,114,507 レビューを有する。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であり、75,519 アイテム、170,645 ユーザー、475,327 レビューを有する。MovieLens は、映画の評価がまとめられており、10,677 アイテム、69,878 ユーザー、10,000,054 レビューを有する。

4. 実験結果

図1のグラフは、横軸に反復回数、縦軸に尤度をプロットした実験結果を示す。(a) アニコレ (b) @cosme(c)

[†]静岡県立大学 経営情報学部

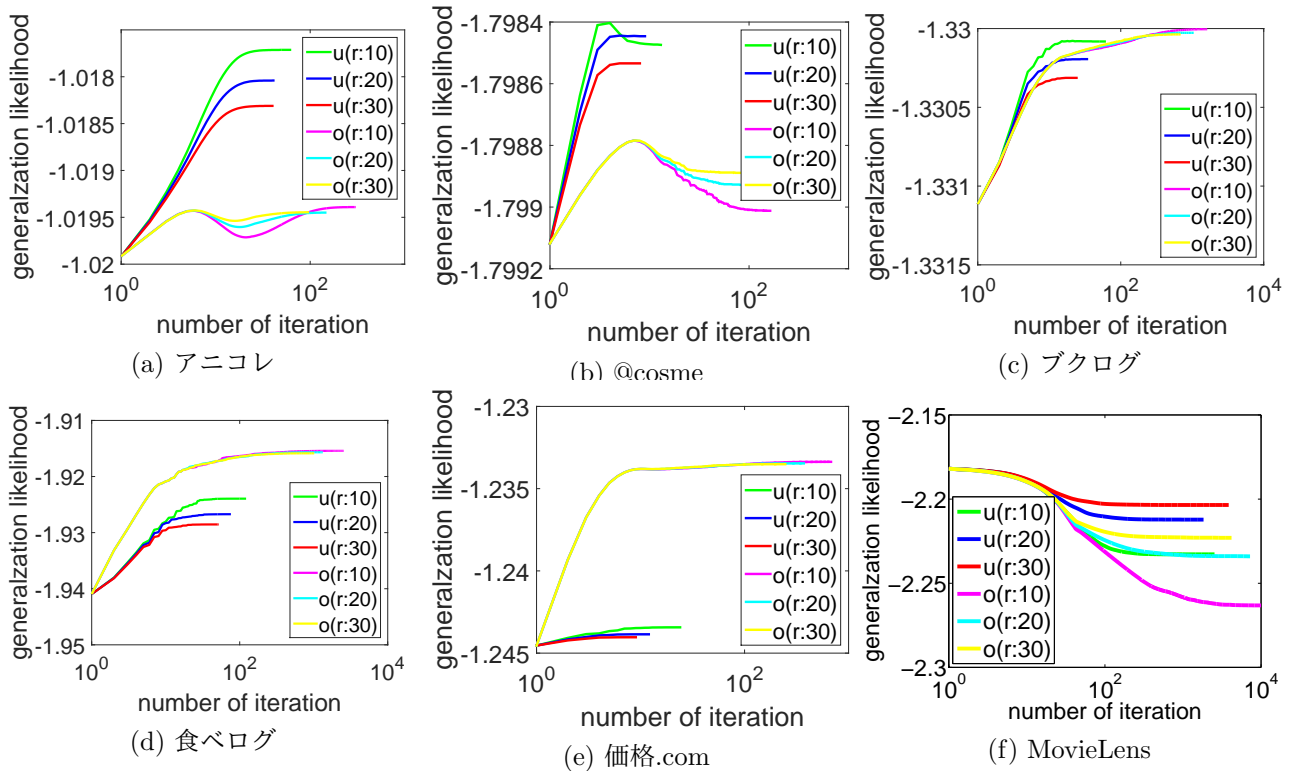


図 1: 反復回数と尤度による各レビューサイトの評価結果

ブクログ (d) 食べログ (e) 価格.com (f) MovieLens の結果を示す。

ユーザー法に重みを付け、尤度の向上が顕著に見られるものが (a) アニコレと (b) @cosme である。つまり、(a) と (b) は誰がレビューしているのかが大事であり、注目すべきである。(c) ブクログはユーザー法、order 法ともに同じような結果になっているが、予想されるのは、影響度の高いユーザーがいち早くレビューしたのではないかと考えられる。(d) 食べログ、(e) 価格.com はユーザーより、レビューの順番に重みを付けることのほうが尤度の向上が見て取れる。(f) MovieLens がユーザー、レビュー順どちらも示さなかった。これは (f) は完全なアンケートでのデータで、他のユーザーの評価は見えない。そのためこの結果は不自然ではないと考えられる。

以上より、提案指標により、(a) アニコレ、(b) @cosme、及び (c) ブクログでは多くのアイテムにレビューし、そのレビューが他のユーザーに影響を与えるようなユーザーの抽出が期待できる。このようなユーザー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、後続ユーザー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、このようなユーザーのレビュー情報の提示が役立つと考えられる。これに対して、(d) 食べログと (e) 価格.com では、ユーザーの影響よりレビューの順番に注目すべきと考えられる。

5. おわりに

本研究では、投票者モデルに基づき、6つのレビューサイトを用いてユーザーのスコアを計算し、レビューサイトにおいて影響力のあるユーザーを分析した。分析結果より、ユーザー影響度は、レビューサイトごとに異なる傾向を示すことがわかった。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.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 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).