

時間減衰ダイナミクスを導入したモデルによる 意見形成とユーザ投稿間隔の分析

木村 唯香[†] 伏見 卓恭^{††} 斉藤 和巳^{†,††}

[†] 静岡県立大学経営情報学部 〒422-8526 静岡県静岡市駿河区谷田 52-1

^{††} 静岡県立大学経営情報イノベーション研究科 〒422-8526 静岡県静岡市駿河区谷田 52-1

E-mail: †{b10031,j11507,k-saito}@u-shizuoka-ken.ac.jp

あらまし ユーザ行動のモデル化に向け、時間減衰ダイナミクスを導入した意見形成モデルが提案され評価されている。しかし、その評価はユーザのレビューを投稿するタイミングや時期を考慮しないものであった。そこで、本稿ではレビューサイトによる時間減衰ダイナミクスを導入したユーザ投稿間隔モデルを提案する。レビューデータとして書籍レビューサイト（ブックログ）と映画レビューサイト（Yahoo!映画）のデータを用い、提案モデルを評価する。また、提案モデルで推定したパラメータの値とアイテムの関係を考察する。さらに、意見形成モデルと投稿間隔モデルの結果を比較分析し、評価する。

キーワード 意見形成モデル、時間減衰、投稿間隔モデル、レビューサイト

Yuika KIMURA[†], Takayasu FUSHIMI^{††}, and Kazumi SAITO^{†,††}

[†] School of Management and Information, University of Shizuoka

52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

^{††} Graduate School of Management and Information of Innovation, University of Shizuoka

52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

E-mail: †{b10031,j11507,k-saito}@u-shizuoka-ken.ac.jp

1. はじめに

SNS やレビューサイトをはじめとするソーシャルメディアの台頭により、Web サービス上でのユーザ間のインタラクションがあらゆる場面で見受けられるようになっている。同様に、ソーシャルネットワーク上でのユーザの意見形成過程 [1] に関する研究が注目を浴びている。

本稿では、レビューサイトにおけるユーザの行動モデルとして [2]、時間減衰ダイナミクスを持つ意見形成モデル [3] に着目する。レビューサイトにおけるユーザのアイテムに対する評点行動を分析、評価する。さらに、時間減衰ダイナミクスを持つ投稿間隔モデルを提案し、上述した意見形成モデルとの比較やモデルのパラメータ値とアイテムの関係性について分析、考察する。

本稿の構成は以下の通りである。最初にユーザの行動モデルとして、時間減衰ダイナミクスを導入した意見形成モデルについて説明し、新たに時間減衰ダイナミクスを導入したレビュー投稿間隔モデルを提案する。3 節でこれらのモデルに対し、オンラインレビューサイトのレビューデータを用いて評価する。

最後に本稿のまとめと今後の展望について述べる。

2. 時間減衰モデル

レビューサイトでは、ユーザがアイテムに対してレビューを投稿する活動が頻繁に行われている。これらのユーザ行動と対象となるアイテムの特性の関係について分析するために、ユーザの意見形成モデルおよびユーザのレビュー投稿間隔モデルについて説明、提案する。アイテム i に関して、ユーザ v が時刻 t で評点 k のレビューをしたというリストの集合 $D(i) = \{\dots, (v, t, k), \dots\}$ を入力とする。

任意の時刻 $t > 0$ とノード（ユーザ）集合 $v \in V$ に対して、時刻 t より前に個人（ノード） v が意見を表明した時刻全体の集合 $M(t, v)$ を考える。また、 $k = 1, \dots, K$ に対して、 $M(t, v)$ の部分集合 $M_k(t, v) = \{\tau \in M(t, v); g_t(v) = k\}$ を考える。ここに $M_k(t, v)$ は、時刻 t より前にノード v が意見 k を表明した時刻全体の集合であり、 $g_t(v)$ は、時刻 t におけるノード v の意見である。

2.1 時間減衰ダイナミクスをもつ意見形成モデル

人々は意見を決定するとき、周囲の人の最新の意見だけで

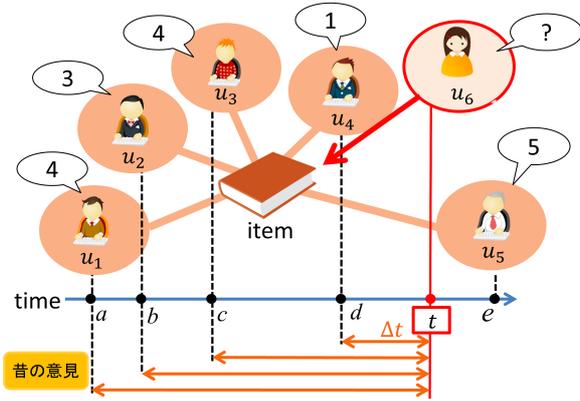


図 1 過去の意見も考慮する意見形成

なく、過去の意見をも考慮すると考えられる．図 1 の例では、ユーザ u_6 は周囲の人が過去に表明した意見を考慮に入れ自身の意見を決定しようとする．このような状況をモデル化した TDV モデル (Temporal Decay Voter Model) について説明する．

過去に表明されたすべての意見を考慮する投票者モデルは、

$$P(g_t(v) = k) = \frac{1 + \sum_{u \in B(v)} |M_k(t, u)|}{K + \sum_{u \in B(v)} |M(t, u)|}, (k = 1, \dots, K) \quad (1)$$

で定義される． $B(v)$ はノード v の親ノード集合である．この投票者モデルを“基礎 TDV モデル”と呼ぶ．

これまで、過去のすべてのオピニオンが等しく重みづけらると仮定した．しかしながら、非常に古いオピニオンはほとんど影響を及ぼさないのが普通であり、一般には、古いオピニオンほど影響が小さいと考えられるため、このような効果を反映させるために、2 つの減衰関数を導入する．1 つ目は指数的減衰関数であり、

$$\rho_{\text{exp}}(\Delta t; \lambda) = \exp(-\lambda \Delta t) \quad (2)$$

で定義される．ここに、 $\lambda \geq 0$ はパラメータ、 $\Delta t = t - \tau$ はオピニオン採用時刻 t とオピニオン表明時刻 τ との時間差を表している．本稿では、この指数的減衰関数を導入した TDV モデルを“指数的 TDV モデル”と呼び以下のように定義する．

$$P_{\text{exp}}(g_t(v) = k) = \frac{1 + \sum_{u \in B(v)} \sum_{s \in M_k(t, u)} \rho_{\text{exp}}(t - s; \lambda)}{K + \sum_{u \in B(v)} \sum_{s \in M(t, u)} \rho_{\text{exp}}(t - s; \lambda)} \quad (3)$$

2 つ目はべき乗的減衰関数であり、

$$\rho_{\text{pow}}(\Delta t; \lambda) = (\Delta t)^{-\lambda} = \exp(-\lambda \log \Delta t) \quad (4)$$

で定義される．ここに、 $\lambda \geq 0$ はパラメータである．本稿では、このべき乗的減衰関数を導入した TDV モデルを“べき乗的 TDV モデル”と呼び以下のように定義する．

$$P_{\text{pow}}(g_t(v) = k) = \frac{1 + \sum_{u \in B(v)} \sum_{s \in M_k(t, u)} \rho_{\text{pow}}(t - s; \lambda)}{K + \sum_{u \in B(v)} \sum_{s \in M(t, u)} \rho_{\text{pow}}(t - s; \lambda)} \quad (5)$$

2.2 時間減衰ダイナミクスをもつ投稿間隔モデル

ユーザによるクチコミやレビューは、アイテムの性質や時期によってその投稿間隔が異なると考えられる．コンスタントにレビューされるロングセラーのようなアイテムもあれば、一時

的に急激にレビューされるヒット商品のようなアイテムも存在する．このような事実を分析するために、本稿では、アイテムごとのレビュー投稿の時系列に関して、TDI モデル (Temporal Decay voting Interval Model) を提案する．

定常性を想定した通常の指数分布に基づく投稿間隔モデルは、

$$P(\delta \tau_t) = \phi \exp(-\phi \delta \tau_t) \quad (6)$$

で定義される．ここで、 $\delta \tau_t$ は時刻 t における投稿間隔を表し、 ϕ は投稿間隔モデルを規定する指数分布のパラメータを表す．意見形成モデルと同様に、指数的減衰関数とべき乗的減衰関数を導入した投稿間隔モデルを考える．以降では、指数的減衰関数を導入した TDI モデルを“指数的 TDI モデル”と呼び以下のように定義する．

$$P_{\text{exp}}(\delta \tau_t) = \phi_{\text{exp}}(\lambda) \exp(-\phi_{\text{exp}}(\lambda) \delta \tau_t) \quad (7)$$

$$\phi_{\text{exp}}(\lambda) = \frac{\sum_{v \in V} \sum_{s \in M(t, v)} \rho_{\text{exp}}(t - s; \lambda)}{\sum_{v \in V} \sum_{s \in M(t, v)} \rho_{\text{exp}}(t - s; \lambda) \delta \tau_s} \quad (8)$$

一方、べき乗的減衰関数を導入した TDI モデルを“べき乗的 TDI モデル”と呼び以下のように定義する．

$$P_{\text{pow}}(\delta \tau_t) = \phi_{\text{pow}}(\lambda) \exp(-\phi_{\text{pow}}(\lambda) \delta \tau_t) \quad (9)$$

$$\phi_{\text{pow}}(\lambda) = \frac{\sum_{v \in V} \sum_{s \in M(t, v)} \rho_{\text{pow}}(t - s; \lambda)}{\sum_{v \in V} \sum_{s \in M(t, v)} \rho_{\text{pow}}(t - s; \lambda) \delta \tau_s} \quad (10)$$

3. 評価実験

3.1 データセット

本稿では、書籍に関するレビューサイトの一つであるブックログ [4] と、映画に関するレビューサイトの一つである Yahoo!映画 [5] のレビュー情報を実験データとして用いた．

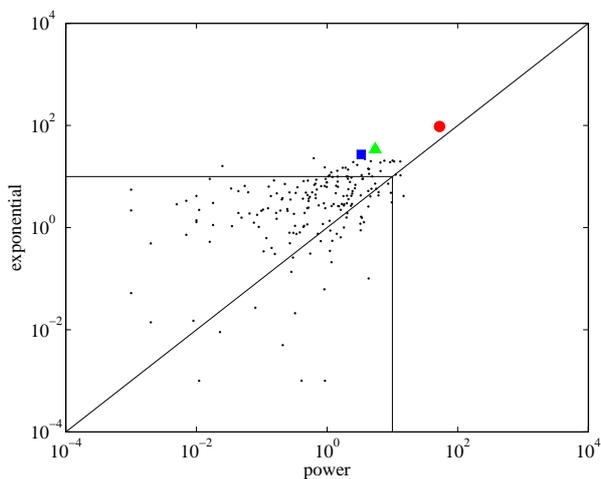
ブックログデータは、2012 年 9 月からおよそ 1 か月間にわたりクローリングし取得したものである．今回使用するデータは、レビュー数が 500 以上のアイテムに絞り、アイテム数は 418、ユーザ数は 74,194、レビュー数は 358,351 である．

映画データは、2012 年 11 月からおよそ 1 か月間にわたりクローリングし取得したものである．今回使用するデータは、レビュー数が 500 以上のアイテムに絞り、アイテム数は 384、ユーザ数は 133,668、レビュー数は 398,767 である．

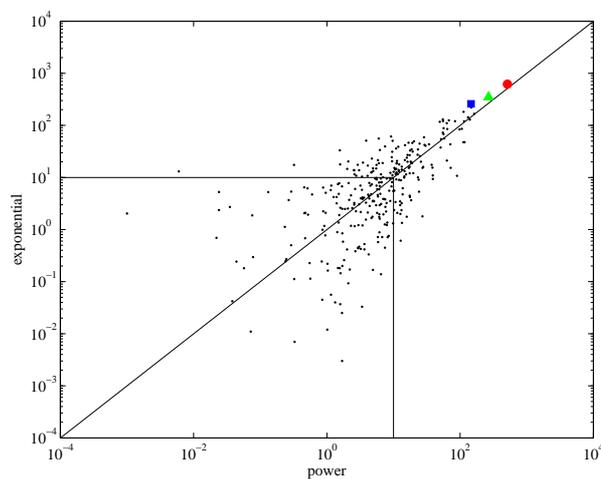
3.2 意見形成モデルの評価

まず、指数的 TDV モデル、べき乗的 TDV モデルのそれぞれに対する基礎 TDV モデルとの尤度比をプロットしたものを図 2 に示す．横軸がべき乗的、縦軸が指数的 TDV モデルとの尤度比である．水平及び垂直な直線はカイ自乗値を表す．すなわち、水平直線より上あるいは垂直な直線より下にあるアイテムは、統計的に有意にそのモデルに従うことが言える．

図 2(a) から、ブックログデータでは多くのアイテムがべき乗的 TDV モデルと比較して指数的 TDV モデルに従う傾向がある．指数的 TDV モデルと基礎 TDV モデルの尤度比の値で降順に並び替えた結果、“もし高校野球の女子マネージャーがドラッカーの『マネジメント』を読んだら”(図中の) や“ス



(a) ブ로그データ



(b) 映画データ

図 2 各モデルに対する尤度比の散佈図 (TDV)

表 1 指数的 TDV モデルにおける推定パラメータランキング
(a) ブ로그データ (b) 映画データ

順位	パラメータ	アイテム名	レビュー数
1	1.41E-02	もしドラ	2973
2	1.19E-02	舟を編む	1559
3	1.10E-02	謎解きはディナーのあとで	2032
-	-	-	-
416	7.45E-09	レインツリーの国 (新潮文庫)	1157
417	7.19E-09	八日目の蝉 (中公文庫)	1162
418	5.83E-09	阪急電車 (幻冬舎文庫)	2563

順位	パラメータ	アイテム名	レビュー数
1	1.03E+00	しんぼる【2009】	1196
2	9.56E-01	クロズド・ノート【2007】	2093
3	4.94E-01	余命 1ヶ月の花嫁【2009】	1766
-	-	-	-
382	7.01E-09	チェンジリング【2008】	1230
383	6.86E-09	ナイト ミュージアム【2006】	1096
384	6.67E-09	パイ・カリ/生命の泉【2011】	1054

テップファザー・ステップ (講談社文庫) (図中の), “夏と花火と私の死体 (集英社文庫) (図中の) などの比較的人気の書籍が含まれていた。

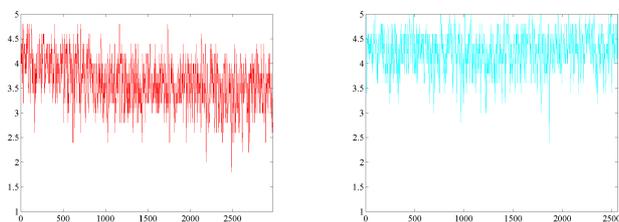
図 2(b) から、映画データでも多くのアイテムがべき乗的 TDV モデルと比較して指数的 TDV モデルに従う傾向がある。指数的 TDV モデルと基礎 TDV モデルの尤度比の値で降順に並び替えた結果、“ALWAYS 三丁目の夕日 (図中の)”や“硫黄島からの手紙 (図中の)”、“ニュームーン/トワイライト・サーガ (図中の)”などの映画が含まれていた。また、ブ로그データに比べて有意水準を大きく超えるアイテムが多い結果となった。

次に、表 1(a) にブ로그データの指数的 TDV モデルにおけるパラメータ λ の値で降順ソートした結果を示す。ランキング上位の書籍は単行本が多く含まれていた。単行本とは、通常文芸誌で発表された作品をきれいにまとめ出版され、後述する文庫本より先行して販売される。従って、単行本の読者はアーリーアダプター (初期採用者) が多く、その書籍に対する世間一般の評判が定着する前に書籍を購読している傾向がある。そのため、書籍に対する評価にバラつきがあると考えられる。

一方、ランキング下位の書籍は文庫本が多く含まれていた。文庫本とは、単行本が出版されてからある程度時期を経て出版されるため、世間一般の評価が定まっているものを読むことになる。また、「あらすじ」と「解説」がついており、興味のない読者は購読しない場合もある。そのため、書籍に対する

評価は比較的安定すると考えられる。

さらに、続編が複数巻に渡って出版されている作品の順位は、その出版時期に従う傾向が見られた。前半巻は上位に位置し、後半巻は下位に位置していた。シリーズもの前半巻と後半巻には、出版時期に差がある。従って、単行本と文庫本との関係と同様に出版時期の影響を受けていると考えられる。シリーズものにおける先行して出版された前半巻の書籍に比べ、後半巻は前半巻で絞り込まれた読者層が集中していると考えられる。そのため、続編を期待する似通った嗜好をもつ読者層がレビューをすることになり、後半巻になるにつれて、書籍に対する評価が安定していくと考えられる。



(a) もしドラ

(b) 阪急電車

図 3 評点時系列 (5 項移動平均)

式 2 より、パラメータ λ の値が大きいくほど、より過去の情報を考慮しないことを表し、値が小さいほど過去の情報を考慮したレビュー時系列データになっていることを表す。このことから、ランキング上位の書籍は、評点の変動が大きく、ほとん

どランダムに評点が付けられていることが示唆される．逆にランキング下位の書籍は，評点の変動が比較的小さいことが示唆される．図 3 は，横軸に時間軸，縦軸に 5 項移動平均の値をプロットしたものである．図 3(a) および (b) より，上位の書籍および下位の書籍の評点の時系列変動をみると，比較的上位の書籍は時間ごとの評点のバラつきが大きい，下位の書籍は連続して近い値の評点が付けられている傾向があることを確認した．

表 1(b) に映画データの指数的 TDV モデルにおけるパラメータ λ の値で降順ソートした結果を示す．ランキング上位の映画は邦画が比較的多く含まれていた．逆に，ランキング下位の映画は洋画が比較的多く含まれていた．邦画は，好き嫌いがあったり，個人の嗜好が偏るため，評点がばらつきやすい．その理由として，洋画に比べて鑑賞者は幅広い年齢層（老若男女）であるためではないかと考えられる．さらに，上位の映画の多くは主演俳優が映画好きでない人にも知られている日本人・有名人であるため，映画を観る動機として，内容重視と出演者重視の違いによってばらつきが出やすいのではないかと考えられる．洋画は，邦画よりも内容重視の傾向が強いと考えられる．邦画よりも鑑賞者の層が絞りこまれており，邦画のような出演者による評価付けはあまりみられない（良い内容は高い／悪い内容は低い）．このような理由から，パラメータランキングの結果，すなわち，アイテムの評点分布の安定具合が決定づけられるのではないかと考察した．

3.3 投稿間隔モデルの評価

意見形成モデルと同様に，指数的 TDI モデル，べき乗的 TDI モデルのそれぞれに対する基礎 TDI モデルとの尤度比をプロットしたものを図 4 に示す．横軸がべき乗的，縦軸が指数的 TDI モデルとの尤度比である．

図 4(a) を見ると，ブクログデータでは多くのアイテムが比較的べき乗 TDI モデルに有意に従う傾向にある．図中の λ は，それぞれ 1 位の“ハチミツとクローバー 10 (クイーンズコミックス コーラス)”，2 位の“KAGEROU”，3 位の“ヘタリア Axis Powers”である．後述する，べき乗 TDI モデルにおけるパラメータの値で降順ソートした結果における上位の書籍である．

図 4(b) を見ると，映画データでも多くのアイテムが比較的べき乗 TDI モデルに有意に従う傾向にある．図中の λ は，それぞれ 1 位の“252 生存者あり”，2 位の“カラフル”，3 位の“BECK”である．後述する，べき乗 TDI モデルにおけるパラメータの値で降順ソートした結果における上位の映画である．またブクログデータと比較して，有意水準を大きく超えるアイテムが多い結果となった．

次に，表 2 にブクログデータでのべき乗的 TDI モデルにおけるパラメータ λ の値で降順ソートした結果の一部を示す．上位，つまりパラメータが大きい書籍は，漫画が多く含まれていた．これは，投稿間隔にバラつきがあることを示している．この理由として，漫画のプロダクトライフサイクルが短いことにあると考える．漫画の中でも，上位はほとんどのタイトルが連載やアニメの放送が終了したタイトルであった．下位は漫画の連載やアニメの放送が続いているタイトル，映画化されたタイ

トルが位置していた．漫画作品はアニメ化や映画化の影響を受けやすいものであること示している．また，連載を継続するなどして，長期的に読者がその作品に触れる機会を設けることで話題性を保ち，衰退期を遅らせているといえる．一方，下位の書籍には小説が多く含まれていた．小説は漫画よりもプロダクトライフサイクルが長いことがいえる．とくに下位に占める小説は，有川浩や伊坂幸太郎，東野圭吾などの有名作家の作品が多くを占めていた．コンスタントにレビューされる話題性のある作家を示している．なお，書籍では，べき乗的 TDI モデルにおけるパラメータ λ の降順ソート結果と，レビューの評点の高さや低さとの相関は見られなかった．

次に，表 3 に映画データでのべき乗的 TDI モデルにおけるパラメータ λ の値で降順ソートした結果の一部を示す．邦画と洋画の 2 つに分類し，比較を行ったが，ランキングにおける分布に特徴はみられなかった．そこで，シリーズもの一作品のみのタイトルの区別を加え，4 つに分類し，分析を行った．特徴が顕著に表れたのが，シリーズものの邦画と一作品のみの洋画であった．シリーズものの邦画は上位を多く占め，一作品のみの洋画は下位を多く占めていた．すべての洋画が日本で公開されるわけではない．ある程度，海外で観客数をあげた人気作が放映されるため，話題作になる可能性が邦画よりも高いと考えられる．また，書籍は絶版になるまで半永久的にその書籍を読む機会があるが，映画は放映されている期間が限られているという理由から，映画は書籍と比較して話題性を保つのが難しいと考えられる．放映期間または，その前後ではレビュー数の上昇がみられるが，しばらくすると，その映画に触れる機会がなくなり，レビュー数は停滞傾向にある．DVD 化やテレビ放映などでレビュー数が再び上昇することもあり，レビューの投稿間隔にばらつきが出やすいと考えられる．

4. おわりに

本稿では，TDV モデルと TDI モデルに着目し，書籍のレビューサイトであるブクログを対象に評価した．意見形成モデルでは，比較的指数的 TDV モデルに従う傾向があることを確認した．さらに，推定したパラメータによるランキングでは，上位に単行本，下位に文庫本が多く並ぶなど，直観にあった結果が得られた．対して，投稿間隔モデルでは，比較的べき乗的 TDI モデルに従う傾向があることを確認した．推定したパラメータによるランキングでは，上位に単行本，下位に文庫本が多く並ぶなど，直観にあった結果が得られた．今後は，書籍のジャンルや出版時期，ドラマ化・映画化の有無などを考慮に入れて分析を進めていきたい．書籍だけでなく，他のアイテムを取り扱うレビューサイトなども焦点に入れて評価していきたい．謝辞 本研究は，科学研究費補助金基盤研究 (C)(No.23500312) の補助を受けた．

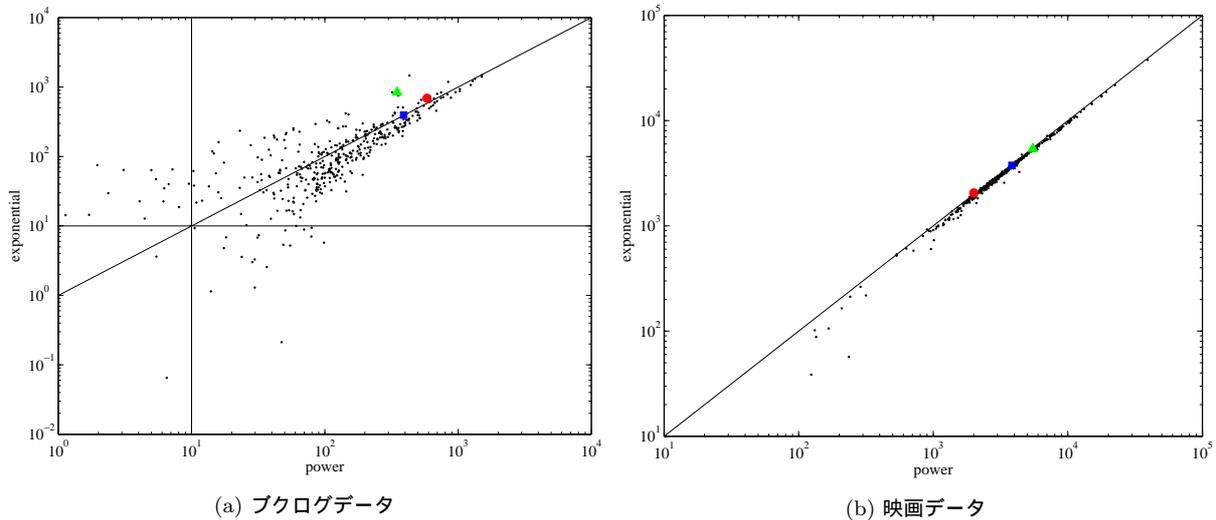


図 4 各モデルに対する尤度比の散布図 (TDI)

表 2 べき乗的 TDI モデルにおける推定パラメータランキング (ブックログデータ)

順位	パラメータ	アイテム名	レビュー数
1	2.39E+00	ハチミツとクローバー 10(クイーンズコミック コーラス)	510
2	2.32E+00	KAGEROU	877
3	1.68E+00	ヘタリア Axis Powers	2032
4	1.63E+00	ホームレス中学生 (665)	665
5	1.57E+00	黒執事 1 (1) (G ファンタジーコミックス)	541
-	-	-	-
414	1.68E+00	ビブリア古書堂の事件手帖 3 栞子さんと消えない絆 (メディアワークス文庫)	547
415	1.11E-01	三匹のおっさん (文春文庫)	563
416	9.76E-02	ビブリア古書堂の事件手帖 2 栞子さんと謎めく日常 (メディアワークス文庫)	876
417	6.10E-05	少女	562
418	3.05E-05	三匹のおっさん	731

表 3 べき乗的 TDI モデルにおける推定パラメータランキング (映画データ)

順位	パラメータ	アイテム名	レビュー数
1	5.00E+00	252 生存者あり【2008】	512
2	4.79E+00	カラフル【2010】	968
3	3.92E+00	BECK【2010】	913
4	3.91E+00	県庁の星【2006】	616
5	3.80E+00	ミッドナイト イーグル【2007】	631
-	-	-	-
380	1.38E+00	きみに読む物語【2004】	589
381	1.33E+00	バイオハザード V:リトリビューション【2012】	560
382	1.26E+00	バック・トゥ・ザ・フューチャー【1985】	728
383	1.26E+00	最強のふたり【2011】	557
384	1.21E+00	ショーシャンクの空に【1994】	1144

文 献

- Databases (ECML PKDD 2012)
- [1] E.Even-Dar, A.Shapira, A Note on Maximizing the Spread of Influence in Social Networks, Lecture Notes in Computer Science, Vol.4858, (2007).
 - [2] M.J.Salganik, P.S.Dodds, and D.J.Watts, Experimental Study of Inequality and Unpredictability in an Art - cial Cultural Market, Science 10, pp.854-856, February 2006.
 - [3] M.Kimura, K.Saito, K.Ohara, and H.Motoda, Opinion formation by voter model with temporal decay dynamics, Proceedings of 2012 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in
 - [4] <http://booklog.jp/>
 - [5] <http://movies.yahoo.co.jp/>