

時間減衰付きカテゴリ選択モデルを用いたレビュー傾向分析

Analysis of Review Tendency using a Category Selection Model with Temporal Decay

伏見 卓恭[†]
Takayasu Fushimi

佐藤 哲司[†]
Tetsuji Satoh

斉藤 和巳[‡]
Kazumi Saito

風間 一洋[§]
Kazuhiro Kazama

1. はじめに

Web 技術の進展により、ユーザ（消費者）は購入したアイテム（商品）に関する評価・感想をレビューサイトへ投稿する機会が増加している。それに伴い、豊富なユーザ嗜好に関する情報を入手可能になり、ユーザの嗜好に応じた推薦サービスが注目を浴びている。一方、多様なユーザがそれぞれのアイテムに対し、さまざまな観点からレビューを投稿することから [1]、新たにアイテムを購入しようとして検討しているユーザがレビューを参考にする際、どんなユーザのレビューに着目すべきかを判断するのは困難な場合がある。この背景を踏まえて、レビューしているユーザがそのアイテムに関して「どの程度詳しいか」や「どのような観点で評価しているか」、「どの程度そのアイテムにこだわりがあるか」などを定量化し表示することは、新たにアイテムを購入する際に有用な情報となりうる、かつ、重要な研究課題である。

本研究ではその第一歩として、過去のレビュー履歴から各ユーザがアイテムのカテゴリにどの程度こだわりがあるかについて着目する。すなわち、ユーザはアイテムを選択する際に、そのアイテムの属性情報を意識的に、あるいは無意識に考慮して選択していると考えられる。さらに、上述の傾向は一生変わらないものではなく一時的な場合もあり、いわゆるマイブームのようなものが存在していると考えられる。そこで、ユーザのアイテム選択行動をカテゴリの多項分布モデルに従うとしたモデルを構築する。さらに、アイテム選択の際のカテゴリ依存度が一過性であるか持続的なものであるかを評価するために、文献 [2] の時間減衰付き意見形成モデルをベースに、時間減衰を導入したカテゴリ選択モデルを提案する。そして、各ユーザが同一カテゴリのアイテムを選択する傾向にあるかどうかをモデルにより評価する。実際のレビューデータを用いた評価実験では、モデルの尤度や減衰パラメータでユーザをランキングし、その特徴や傾向について分析する。

2. 関連研究

この章では、ユーザのアイテム選択に関する関連研究について触れる。ユーザがある特定のブランドを繰り返し購入することを、マーケティング用語でブランド・ロイヤルティと呼ぶ。本研究のカテゴリ選択も、ブランド・ロイヤルティに類似する概念である。文献 [3] では、各ユーザのブランドに対する「購買比率尺度」、「連続購買尺度」、「購買確率尺度」などの尺度を整理し

ている。これらは、購入アイテムのうち特定のブランドを選択している比率や、特定ブランドアイテムの連続購買数の平均などにより定量化され、マーケティング分野では重要視されていた。

文献 [4] では、EC サイトにおけるユーザの閲覧行動および購買行動の流れに着目し、購入検討の候補となったアイテムに共通する商品選択条件語 (ISRW) を推定する手法を提案した。評価実験により、TF・IDF や RAKE などの既存手法より高い推定精度が出ること示し、個人ごと、あるいは、閲覧行動の目的ごとに ISRW が異なることを実証した。この研究では、アイテムの閲覧行動と購買行動を用いているが、本研究ではアイテム購買後のレビュー行動、および、アイテムの属性情報に着目している点で異なる。

文献 [5] では、あるキーワードが時間軸方向に密に出現するかを否かを判定するバースト度を提案した。この手法を用いれば、大量のデータから一時的に特定のカテゴリのアイテムがバーストするような場合を検出できるが、各ユーザのレビューデータなど比較的小規模なストリームデータからのバースト検出は困難な場合がある。一方、文献 [2] では、ユーザがネットワーク上の隣接ノードの最新の意見を考慮する単純 Voter Model を拡張し、過去に有していた意見をも考慮するモデルを提案した。さらに、時間減衰を導入することにより時間差の大きい意見を割り引いて考慮するモデルも提案した。評価実験により、古い意見まで参考にするユーザもいたり、逆に古い意見はあまり考慮しないユーザもいるなど、時間減衰の効果を検証した。このモデルを土台とし、時間減衰を導入したカテゴリ選択モデルを提案する。

3. 提案手法

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、ユーザがカテゴリを考慮してアイテムを選択するモデルを構築する。

3.1. カテゴリ選択モデル

ユーザ u に関して、時刻 t にアイテム i にレビューをしたというリストの集合 $D_u = \{\dots, (i, t), \dots\}$ を入力とする。ユーザ u が時刻 t より前にレビューした時刻の集合を $R_u(t)$ とあらわす。つまり、 $t > \tau \in R_u(t)$ である。このうち、レビュー対象のアイテムカテゴリが c であるものを $R_u^{(c)}(t) \subset R_u(t)$ と表記する。アイテム i のカテゴリを $c(i)$ 、総カテゴリ数を C とするとき、ユーザ u が時刻 t にカテゴリ c に属するアイテムのいずれかをレビューする確率を次式により計算

[†]筑波大学図書館情報メディア系

[‡]静岡県立大学経営情報学部

[§]和歌山大学システム工学部

する。

$$P_u(c, t) = \frac{1 + |R_u^{(c)}(t)|}{C + |R_u(t)|} \quad (1)$$

このカテゴリ選択モデルにおいて、ユーザは今までにレビューしたアイテムカテゴリの多項分布に従って、新たなアイテム(カテゴリ)を選択する。すなわち、ユーザのアイテム選択は対象アイテムのカテゴリに依存するというモデルである。本稿では、基本カテゴリ選択モデルと呼ぶ。

3.2. 時間減衰モデル

ユーザがアイテムをカテゴリにより選択するという行動は、一時的な場合もあれば、長期的な場合も考えられる。この点を考慮するために、上述のモデルに時間減衰を導入する。すなわち、多項分布型のカテゴリ選択モデル(1)が長期間にわたり適用できるか、短期間のものなのかをパラメータ λ により制御する。時間減衰関数としてパラメータ $\lambda \geq 0$ をもつ指数関数型減衰関数 $\rho(\Delta t; \lambda) = \exp(-\lambda \Delta t)$ を用いる。時間減衰を導入したカテゴリ選択モデルは以下ようになる。

$$P_u(i, t; \lambda) = \frac{1 + \sum_{\tau \in R_u^{(c(i))}(t)} \rho(t - \tau; \lambda)}{C + \sum_{\tau \in R_u(t)} \rho(t - \tau; \lambda)} \quad (2)$$

本稿では、時間減衰付きカテゴリ選択モデルと呼ぶ。このモデルは、時刻 t と過去のレビュー時刻 τ が離れている場合は、近い場合より割り引いて考慮することを意味する。すなわち、パラメータ λ が大きければ、時間減衰が強きはたらき、ユーザが古くレビューしたアイテムのカテゴリを割り引いて考慮する。逆に、 λ の値が小さければ、時間減衰の効果は弱く、ユーザが古くレビューしたアイテムのカテゴリも十分考慮していることになる。パラメータ $\lambda = 0$ のとき、すべての重みは1となり、レビューされた時刻を考慮しない基本カテゴリ選択モデルとなる。

3.3. 解法アルゴリズム

時間減衰を導入したカテゴリ選択モデルのパラメータを求める。モデル(2)より、尤度関数は

$$\begin{aligned} \mathcal{L}(D_u) &= \log \left(\prod_{(i,t) \in D_u} P_u(i, t; \lambda) \right) \\ &= \sum_{(i,t) \in D_u} \log \left(1 + \sum_{\tau \in R_u^{(c(i))}(t)} \exp(-\lambda(t - \tau)) \right) \\ &\quad - \sum_{(i,t) \in D_u} \left(C + \sum_{\tau \in R_u(t)} \exp(-\lambda(t - \tau)) \right) \end{aligned} \quad (3)$$

となる。この尤度関数を最大にするようなパラメータ $\hat{\lambda}$ を求める。

λ の現在の推定値を $\bar{\lambda}$ とし、2つの事後確率

$$q_{u,i,t}(\tau; \lambda) = \frac{\exp(-\lambda(t - \tau))}{1 + \sum_{\tau' \in R_u^{(c(i))}(t)} \exp(-\lambda(t - \tau'))}$$

と、

$$r_{u,t}(\tau; \lambda) = \frac{\exp(-\lambda(t - \tau))}{C + \sum_{\tau' \in R_u(t)} \exp(-\lambda(t - \tau'))}$$

を用いて、式(3)をEMアルゴリズムと同様に

$$\mathcal{L}(D_u; \lambda) = \mathcal{Q}(\lambda; \bar{\lambda}) - \mathcal{H}(\lambda; \bar{\lambda})$$

と変形できる。ただし、

$$\begin{aligned} \mathcal{Q}(\lambda; \bar{\lambda}) &= - \sum_{\tau \in R_u^{(c(i))}(t)} q_{u,i,t}(\tau; \bar{\lambda}) \lambda(t - \tau) \\ &\quad + \sum_{\tau \in R_u(t)} r_{u,t}(\tau; \bar{\lambda}) \lambda(t - \tau) \end{aligned}$$

である。 $\mathcal{H}(\lambda; \bar{\lambda})$ は、最大化に直接関係しないため割愛する。ニュートン法を用いて $\mathcal{Q}(\lambda; \bar{\lambda})$ を最大化するようにパラメータ λ を求める。

4. 評価実験

4.1. データセット

本稿では、映画に関するレビューサイトの一つであるYahoo!映画[¶]のレビュー情報を実験データとして用いた。映画レビューデータは、2014年1月にクロールし取得したものであり、最古のレビューは2005年11月、最新のレビューは2014年1月である。今回使用するデータは、アイテム数が39,681、ユーザ数が15,593、レビュー数が719,227である。カテゴリとして、「主演俳優・声優」、「監督」、「ジャンル」を用い、それぞれカテゴリ数は14,969, 13,731, 16である。「ジャンル」は、1つの映画に複数のカテゴリが割り当てられている。

4.2. 実験結果

表1(a),(b),(c)に、カテゴリを「主演俳優・声優」、「監督」、「ジャンル」としたときの、基本カテゴリ選択モデルの対数尤度のランキングを示す。ユーザ名は一部マスキングしてある。表1(a)において上位ユーザの傾向として、ジャッキー・チェン、クリント・イーストウッド、ジョニー・デップなどが主演を務める作品のみを見ているユーザが多く見受けられる。他にも、高山みなみ、松本梨香、田中真弓などが声優を務めるアニメシリーズ映画のみに特化して見ているユーザも多い。一方ランキング下位ユーザの傾向として、洋画や邦画など様々な種類の映画、あるいは、洋画のうち、様々な主演俳優の映画を閲覧しており、「主演俳優・声優」というカテゴリに関してこだわりなく、広く浅く映画を選択していると考えられる。

表1(b)において上位ユーザの傾向として、井筒和幸の作品のみを閲覧するユーザや、成瀬巳喜男、吉田喜重、増村保造、川島雄三、マキノ正博など、同世代の監督の作品を網羅的に閲覧するユーザ、宮崎駿、森田宏幸、近藤喜文、高畑勲など関連する監督の作品を見るユーザなどが見受けられる。一方ランキング下位ユーザの傾向として、一つとして同一監督の作品を閲

[¶]<http://movies.yahoo.co.jp/>

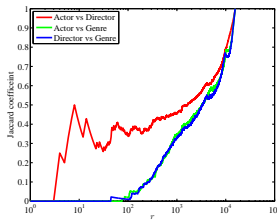


図 1: 尤度ランキングの
相関関係

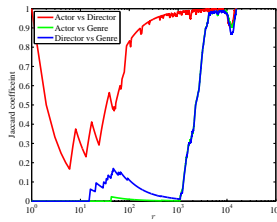


図 2: パラメータランキ
ングの相関関係

覧していないなど、ランダムに選択したのに近いようなユーザが多い。

表 1(c) において上位ユーザの傾向として、閲覧しているほとんど全ての作品にサスペンス、ホラーというジャンルが付されているケース、ドラマ、ラブストーリーの作品、SF・ファンタジー、アクション・アドベンチャーの作品を比較的多めに閲覧しているケースなどが多い。一方ランキング下位ユーザの傾向として、比較的ジャンルに依存せず作品を選択しているパターンが多い。

また、これら 3 つのランキングの相関関係を調べるために、ランキング A における順位 r 位までのユーザ集合を $A(r)$ とし、以下のように Jaccard 係数を計算する。

$$J(A(r), B(r)) = |A(r) \cap B(r)| / |A(r) \cup B(r)| \quad (4)$$

図 1 に、横軸に順位 r 、縦軸に Jaccard 係数をプロットした。「主演俳優・声優」と「監督」の間に幾分か相関関係が見られるが、他のカテゴリペア間には相関があまりないことが伺える。すなわち、「ジャンル」にはこだわりがあるが、「主演俳優・声優」や「監督」にはこだわりがないというユーザが存在することがわかる。

これらのことから、映画を選択する際に、意識的に、あるいは、無意識にカテゴリを考慮しているユーザが存在することが示唆された。また、カテゴリを強く考慮して映画を選択しているユーザとそうでないユーザを分類できれば、専門的観点と新鮮さの観点からレビュー内容を検討することにもつながると考えられる。

次に、時間減衰付きカテゴリ選択モデルの尤度最大にならしめるパラメータ λ のランキング結果を表 2(a),(b),(c) に示す。ただし、対象を表 1 の尤度ランキング上位 1000 ユーザに絞った。表 2(a),(b) において上位ユーザの傾向として、一時的に、同一監督や同一俳優の作品を連続的に閲覧しているユーザが見受けられる。その閲覧時間間隔も比較的短く、かつ、長期的に同一監督、俳優の作品を閲覧していないため、いわゆるマイブームの状態であり、レンタルビデオなどで集中的に閲覧したと考えられる。一方下位のユーザでは、同一監督、俳優の作品を集中的、連続的に閲覧している傾向は少ない。反対に、最古のレビューから最新のレビューまでの間に、同一監督、俳優の作品が比較的多めに出現するようなユーザが多い。このことから、マイブームの状態ではなく、好みの監督、俳優の作品がリリースされたときに閲覧したと考えられる。

表 2(c) において上位ユーザの傾向として、初期の頃はドラマやラブストーリーを閲覧する比率が多かったが、時代が進むにつれアクション・アドベンチャーやサスペンスといったジャンルの作品を閲覧する比率が増えていくなど、閲覧するジャンルの偏りが変遷する傾向にあるユーザが多かった。一方下位のユーザでは、異なるジャンル作品もたまたま閲覧するが、時代が進んでもある一定のジャンル作品を閲覧しているユーザが多かった。

また、これら 3 つのランキングの相関関係を調べるために、図 2 に Jaccard 係数をプロットした。尤度ランキングと同様に、「主演俳優・声優」と「監督」の間に幾分か相関関係が見られるが、他のカテゴリペア間には相関はあまりないことが伺える。「ジャンル」の構成割合は、時代とともに大きく変化することはないが、「出演俳優」や「監督」はそれぞれ流行り・廃りがあったり、あるいは監督が出演俳優を選択する作品もあるため、これらのカテゴリ選択にある程度の相関が生じたと考えられる。これらの結果から、複数のカテゴリにおいてパラメータの値でユーザを定量化することにより、ユーザがどの程度カテゴリにこだわりがあるか、どのような間隔、周期でそのカテゴリの作品を閲覧しているかなどの動向を知ることができる。そのため、他のユーザがレビューを参考にする際の手助けとなることが示唆される。

5. おわりに

本研究では、ユーザのアイテム選択におけるカテゴリ依存度を考慮したモデルを構築し、そのパラメータでランキングすることで、ユーザごとの傾向を分析する手法を提案した。今後は、多様なレビューデータの多様なカテゴリ情報を用いて評価をすすめていきたい。

謝辞 本研究は、JSPS 科研費 (No.15J00735) の助成を受けたものである。

参考文献

- [1] 小西卓哉ほか：周辺文を考慮するトピックモデルを用いた評価側面の推定，第 4 回データ工学と情報マネジメントに関するフォーラム (2012)。
- [2] Kimura, M. et al.: Opinion Formation by Voter Model with Temporal Decay Dynamics, *Proc. of Machine Learning and Knowledge Discovery in Databases*, pp. 565–580 (2012)。
- [3] Jacoby, J. et al.: *Brand Loyalty: Measurement and Management*, A Ronald Press publication, Wiley (1978)。
- [4] 江崎健司ほか：ネットショッピングにおける商品選択条件の推定，第 4 回データ工学と情報マネジメントに関するフォーラム (2012)。
- [5] Kleinberg, J.: Bursty and Hierarchical Structure in Streams, *Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 91–101 (2002)。

表1: 尤度ランキング

(a) 主演俳優・声優			(b) 監督			(c) ジャンル		
順位	対数尤度	ユーザ名	順位	対数尤度	ユーザ名	順位	対数尤度	ユーザ名
1	-7.453	jacki***	1	-7.793	jsjdh***	1	-0.426	jsstr***
2	-7.754	avoos***	2	-8.182	おーる***	2	-0.490	mocos***
3	-7.810	world***	3	-8.424	yota_***	3	-0.493	konoh***
4	-7.910	liklu***	4	-8.474	avoos***	4	-0.508	mkkis***
5	-7.919	avbib***	5	-8.494	kenta***	5	-0.632	fwhz1***
6	-8.103	harim***	6	-8.534	world***	6	-0.636	kobhe***
7	-8.115	yota_***	7	-8.543	agmtk***	7	-0.645	月光の***
8	-8.267	momo_***	8	-8.552	harim***	8	-0.646	black***
9	-8.333	dmw_t***	9	-8.629	文字読***	9	-0.654	kawar***
10	-8.361	yoshi***	10	-8.644	a7113***	10	-0.660	oakka***
:	:	:	:	:	:	:	:	:
15584	-9.615	FU-*****	15584	-9.529	ばより***	15584	-2.444	anara***
15585	-9.615	yokuh***	15585	-9.529	sakur***	15585	-2.444	jaokg***
15586	-9.615	大人失***	15586	-9.529	映画評***	15586	-2.455	akiko***
15587	-9.615	mark6***	15587	-9.529	patie***	15587	-2.462	yrmsc***
15588	-9.615	since***	15588	-9.529	myo_0***	15588	-2.472	afyue***
15589	-9.615	koya5***	15589	-9.529	kyo-i***	15589	-2.472	eight***
15590	-9.615	horro***	15590	-9.530	bu_su***	15590	-2.481	sprem***
15591	-9.615	zymvt***	15591	-9.530	mexic***	15591	-2.531	husse***
15592	-9.615	yosy***	15592	-9.530	コタロ***	15592	-2.542	babib***
15593	-9.616	hello***	15593	-9.530	leon****	15593	-2.591	woode***

表2: パラメータランキング

(a) 主演俳優・声優			(b) 監督			(c) ジャンル		
順位	パラメータ	ユーザ名	順位	パラメータ	ユーザ名	順位	パラメータ	ユーザ名
1	6.595	the_r***	1	4.066	the_r***	1	1.557	kaz69***
2	1.662	maika***	2	1.545	たかち***	2	1.530	映画生***
3	1.649	hetak***	3	1.486	Kuros***	3	1.484	たか***
4	1.632	my_he***	4	1.195	jlynn***	4	1.481	porep***
5	1.628	mtwy***	5	1.195	shish***	5	1.479	jonat***
6	1.622	momot***	6	1.195	tpaga***	6	1.465	シナモ***
7	1.616	nobor***	7	1.194	chi_c***	7	1.458	yomok***
8	1.601	jlynn***	8	1.194	dmw_t***	8	1.455	fry_t***
9	1.601	下医人***	9	1.194	bully***	9	1.450	rokum***
10	1.601	下医悩***	10	1.194	bookb***	10	1.433	akao0***
:	:	:	:	:	:	:	:	:
991	1.059	nekot***	991	0.979	karat***	991	1.078	試写会***
992	1.059	byuju***	992	0.977	fcc5***	992	1.077	go920***
993	1.017	hitsa***	993	0.977	iry4***	993	1.075	green***
994	0.974	映楽亭***	994	0.976	野暮な***	994	1.073	medus***
995	0.974	hi_ta***	995	0.974	mai_k***	995	1.073	ekoek***
996	0.973	sound***	996	0.972	ken90***	996	1.071	えこう***
997	0.920	sizin***	997	0.971	korin***	997	1.068	jmgta***
998	0.861	freem***	998	0.940	tama2***	998	1.067	kaeka***
999	0.707	y_k77***	999	0.930	porep***	999	1.065	mikot***
1000	0.613	osugi***	1000	0.913	sean****	1000	1.061	aoium***