

# 評点時系列データの区間分割法 An Interval Dividing Method for Score Time Series Data

山岸 祐己<sup>♡</sup> 齊藤 和巳<sup>◇</sup> 武藤 伸明<sup>◇</sup>

Yuki YAMAGISHI Kazumi SAITO  
Nobuaki MUTOH

本稿では、評点時系列データの区間分割を主題とする。オンラインレビューサイトでは、あるユーザがあるアイテムにレビューをすることにより、そのアイテムの今後のレビュー傾向が大きく変化することが起こり得る。このような時系列的イベントを検出することは、ソーシャルメディアの研究において重要な課題であると言える。我々は、ユーザの基本評点行動として多項分布モデルを仮定し、尤度比検定の枠組みで、レビュー評点時系列データの区間分割問題を定式化する。さらに、この問題を高速で解くことを目的とした貪欲法に基づく解法と、解の改善を目的とした局所改善法に基づく解法を提案する。実験では、大規模レビュー評点時系列データを用い、提案法によって得られた分割点の有用性を確かめる。

**The word of mouth information in online review sites is affecting various activities from person to person. At review sites, it may occur that the future review tendency of an item changes a lot by few reviews which were rated and posted by users. Namely, it is very important in research of social media to detect those influential reviews or time series events. Thus, we propose an interval dividing method for review-score time series data of online review sites. We assume the user's basic scoring behavior to be a multinomial distribution model, and formulate an interval dividing problem. In the experiments, we use a real big review data and analyze the dividing points that obtained by our proposed method.**

## 1. はじめに

近年、オンラインレビューサイトにおけるユーザのレビュー行動が非常に活発であり、日々大量に投稿されるレビュー情報は、購買行動を始めとしたユーザの多様な活動に影響を及ぼしている。これは、レビューサイトそのものがメディアとして、商品やサービスのプロモーションを左右する重要な情報発信源になりつつあることを示唆している。

<sup>♡</sup> 学生会員 静岡県立大学大学院経営情報イノベーション研究科修士課程

[j12115@u-shizuoka-ken.ac.jp](mailto:j12115@u-shizuoka-ken.ac.jp)

<sup>◇</sup> 正会員 静岡県立大学経営情報学部

[{k-saito,muto}@u-shizuoka-ken.ac.jp](mailto:{k-saito,muto}@u-shizuoka-ken.ac.jp)

しかし、大手のレビューサイトは、レビューそのものは大量に保有しているものの、それらをユーザに情報として上手く発信できているとは限らない。何故なら、常に情報に対して受け身であるライトユーザなどは、「被レビュー数が多い人気商品」や「投稿数が多い有名ユーザのレビュー」のような情報さえ得られれば、それで満足してしまうように思えるからである。現に、Salganikらのレビューサイトに関する大規模な実験 [1] では、アイテムの質そのものよりも、多数のユーザによって生成された社会的情報の方が圧倒的に影響力が強く、ユーザ本来の嗜好すらも捻じ曲げる恐れがあるという結果が出ている。つまり、サイト利用者の大半を占めているであろうライトユーザは、本来自分が欲しているような情報よりも、レビュー数という社会的情報によって信頼性が保証されている情報を優先して得ている可能性が高いということである。これは、大量に存在するレビューのうちのほんの一部しか情報として活かされていないこと、さらに、その他の多くのレビューが「数」という情報に落とし込まれてしまっていることを意味している。かと言って、大量のレビューの中から自分にとって有益な情報を的確に導き出そうとすると、多くの時間と手間を要してしまうため、社会的情報に頼るのは至極当然と言える。

アイテムに対する社会的情報として、広く用いられているのは「平均評点」である。確かに、平均評点はアイテムに対する評価の収束結果として非常に分かりやすく、レビュー数が多ければ多いほどその信頼性が保証されるのも明確である。しかし、その平均評点に至るまで、評点の付き方が異常なまでに変化していた場合は、一概に信頼性が高いとは言いきれない。中でも代表的な例は「サクラ」や「やらせ」といった、ユーザによる意図的な評価の操作で、これらに関しては国内のニュースでも取り上げられる程問題となっている。さらに、意図的でなくとも、他のメディアから発せられた情報によって評点の付き方が変化する様は安易に観測することができ、その理由も商品の改良から流行の変遷まで多種多様である。少なからず起きているこれらのイベントは、各アイテムにおいて重要な情報であるはずだが、平均評点と直近のレビューから読み取ることは難儀である。それどころか、先立った情報が無ければ、地道に一つずつレビューを辿っていても誰もが気付けるわけではない。よって、膨大なレビューの中からこれらのイベントを検出し、アイテムに関する情報として利用できるようにすることは、Web情報学において非常に重要な研究と言える。我々は、SwanとAllan [2] や Kleinberg [3] と同様に、回顧的 (Retrospective) な立場で評点時系列データの区間分割法を提案し、現実の大規模データを用いてイベントの検出を試みる。

## 2. 問題設定

レビュー評点を  $s_n$ 、レビューが投稿された時刻を  $t_n$  とし、評点時系列データを以下のように表す。

$$D = \{(s_1, t_1), \dots, (s_N, t_N)\}. \quad (1)$$

ここで各評点は、1 から  $J$  の整数値で与えられるとする。即ち、 $s_n \in \{1, \dots, J\}$  となる。モデル記述の都合上、各評点  $s_n$  を以下のように  $J$ -次元ベクトルとしてダミー変数を導入

する.

$$s_{n,j} = \begin{cases} 1 & \text{if } s_n = j; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

いま, 多項分布モデルを仮定し, 評点  $j$  が与えられる確率を  $p_j$  として,  $\mathbf{p} = \{p_1, \dots, p_J\}$  のように表せば, 評点の時系列データの対数尤度関数は次式となる.

$$\mathcal{L}(\mathcal{D}; \mathbf{p}) = \sum_{n=1}^N \sum_{j=1}^J s_{n,j} \log p_j. \quad (3)$$

この尤度関数に対して最尤推定量を求めれば, 以下となる.

$$\hat{p}_j = \frac{\sum_{n=1}^N s_{n,j}}{N}. \quad (4)$$

いま,  $K$  個の時刻から構成される分割点集合を  $\mathbf{C}_K = \{T_1, \dots, T_K\}$  とし, 便宜上  $T_0 = t_1$  かつ  $T_{K+1} = t_N$  と設定しておく. また,  $T_{k-1} < T_k$  であるとし,  $\mathbf{C}_K$  による  $\mathcal{D}$  の分割を  $\mathcal{D}_k = \{(s_n, t_n); T_{k-1} < t_n \leq T_k\}$  で定義する. すなわち,  $\mathcal{D} = \mathcal{D}_1 \cup \dots \cup \mathcal{D}_{K+1}$  となり,  $|\mathcal{D}_k|$  は区間  $(T_{k-1}, T_k]$  に含まれる観測時刻数を表す. ここで, 任意の  $k \in \{1, \dots, K+1\}$  に対して,  $|\mathcal{D}_k| \neq 0$  とする. 一方, 対数尤度関数で用いるパラメータ  $\mathbf{p}$  を, 区間  $\mathcal{D}_k$  毎に対応させて  $\mathbf{P}_{K+1} = \{\mathbf{p}_1, \dots, \mathbf{p}_{K+1}\}$  で定義すれば, 分割点集合  $\mathbf{C}_K$  が与えられたときの観測データ  $\mathcal{D}$  に対する対数尤度は次式で計算できる.

$$\mathcal{L}(\mathcal{D}; \mathbf{P}_{K+1}, \mathbf{C}_K) = \sum_{k=1}^{K+1} \mathcal{L}(\mathcal{D}_k; \mathbf{p}_k). \quad (5)$$

よって, 式 (5) の尤度を最大にするパラメータの最尤推定値を  $\hat{\mathbf{P}}_{K+1}$  とすれば, 我々の区間分割問題は,  $\mathcal{L}(\mathcal{D}; \hat{\mathbf{P}}_{K+1}, \mathbf{C}_K)$  を最大化する分割点集合  $\mathbf{C}_K$  を求める問題となる. ただし, 分割点集合  $\mathbf{C}_K$  の導入効果を直接的に評価するため, この問題の別表現として, 尤度比検定の目的関数として我々の区間分割問題を定式化する. つまり, 分割点が  $K$  個存在するとしたときと, 存在しないとしたとき, 即ち  $\mathbf{C}_0 = \emptyset$  のときの尤度比の対数を次式で定義する.

$$\mathcal{LR}(\mathbf{C}_K) = \mathcal{L}(\mathcal{D}; \hat{\mathbf{P}}_{K+1}, \mathbf{C}_K) - \mathcal{L}(\mathcal{D}; \hat{\mathbf{P}}_1, \mathbf{C}_0). \quad (6)$$

本稿では, 式 (6) で定義した  $\mathcal{LR}(\mathbf{C}_K)$  を最大化する分割点集合  $\mathbf{C}_K$  を求める問題を考える.

### 3. 解法

式 (6) の最大化問題を網羅的に解くと, 最適解を得ることが保証されるが, 計算量は  $O(N^K)$  となるため, 観測点数  $N$  が十分に大きくなると, 実用的な時間で結果が得られるのは  $K=2$  程度までである. よって, 任意の  $K$  に対して高速に問題を解くための手法として, 貪欲法に基づくシンプル法を提案する. しかし, シンプル法は比較的プアーな局所解にトラップされるケースが危惧されるため, 局所改善法に基づくムーブ法によって解品質の改善を図る.

#### 3.1 シンプル法

既に選定した  $(k-1)$  個の分割点集合  $\mathbf{C}_{k-1}$  を固定し, 新たに付加するとして最適な分割点  $T_k$  を求め  $\mathbf{C}_{k-1}$  に加えることを,  $k=1$  から  $K$  まで繰り返す. 一般に,  $N$  が十分に大きいとき,  $2(\mathcal{LR}(\mathbf{C}_k) - \mathcal{LR}(\mathbf{C}_{k-1}))$  は  $\chi^2$  分布になることが知られているため, 終了条件に  $\chi^2$  検定を用いる. すなわち, そのアルゴリズムは以下となる.

**S1.**  $k=1, \mathbf{C}_0 = \emptyset$  と初期化する.

**S2.**  $T_k = \arg \max_{t_n} \{\mathcal{LR}(\mathbf{C}_{k-1} \cup \{t_n\})\}$  を求める.

**S3.**  $\mathbf{C}_k = \mathbf{C}_{k-1} \cup \{T_k\}$  と更新する.

**S4.**  $2(\mathcal{LR}(\mathbf{C}_k) - \mathcal{LR}(\mathbf{C}_{k-1}))$  が設定した  $\chi^2$  の棄却点に満たなければ  $\mathbf{C}_{k-1}$  を  $\mathbf{C}_K$  として出力し終了.

**S5.**  $k=k+1$  とし, **S2** へ戻る.

ただし, 分割点集合  $\mathbf{C}_k$  の要素は  $T_{i-1} < T_i$  となるようインデックスを更新するとする. ここで,  $i=2, \dots, k$  である. 明らかに, シンプル法の計算量は  $O(NK)$  となる.

#### 3.2 ムーブ法

既に選定した  $K$  個の分割点集合  $\mathbf{C}_K$  から分割点  $T_k$  を選び, それ以外の分割点集合  $\mathbf{C}_K \setminus \{T_k\}$  を固定し, より望ましい別の分割点  $T'_k$  に置き換えることを繰り返す. ここで,  $\setminus$  は集合差を表す. 明らかに,  $k=1, \dots, K$  のすべてで置き換えできなければ, すなわち, どの  $k$  でも  $T'_k = T_k$  となれば, この方法でさらなる改善はできないので反復を終了させるとする. すなわち, そのアルゴリズムは以下となる.

**S1.** 既に求めた  $\mathbf{C}_K$  を入力とし,  $k=1, h=0$  と初期化する.

**S2.**  $T'_k = \arg \max_{t_n} \{\mathcal{LR}(\mathbf{C}_K \setminus \{T_k\} \cup \{t_n\})\}$  を求める.

**S3.**  $T'_k = T_k$  なら  $h=h+1$ , さもなければ  $h=0$  とし,  $\mathbf{C}_K = \mathbf{C}_K \setminus \{T_k\} \cup \{T'_k\}$  と更新する.

**S4.**  $h=K$  なら  $\mathbf{C}_K$  を出力し終了.

**S5.**  $k=k+1$  とし, **S2** へ戻る.

明らかに, ムーブ法の計算量は  $O(NK)$  以上となる.

### 4. 評価実験

#### 4.1 データセット

実験で使用したデータセットは, @cosme<sup>1</sup> のレビューデータである. @cosme は, 株式会社アイスタイル<sup>2</sup> が運営する日本最大級の化粧品レビューサイトであり, 1999年12月にサービスが開始された. このデータセットは, 2013年8月に@cosmeをクロールして取得したものであり, 443853 ユーザ, 6173 アイテム, 4376241 レビューを有する. なお, クロール対象は被レビュー数  $N$  が 200 以上のアイテムとした. レビューの評点は 0~7 の整数値を取りうるため, 実験時には 1 ずつ加算して 1~8 ( $J=8$ ) として扱う. 図 1 にレビュー評点の分布を示す.

ここで, このデータセットにおけるレビューは, 人間の行動特性に基づいて投稿されたものかどうかを検証する. 人間の行動特性として, 手紙が手元に届いてから返信する

<sup>1</sup><http://www.cosme.net>

<sup>2</sup><http://www.istyle.co.jp/>

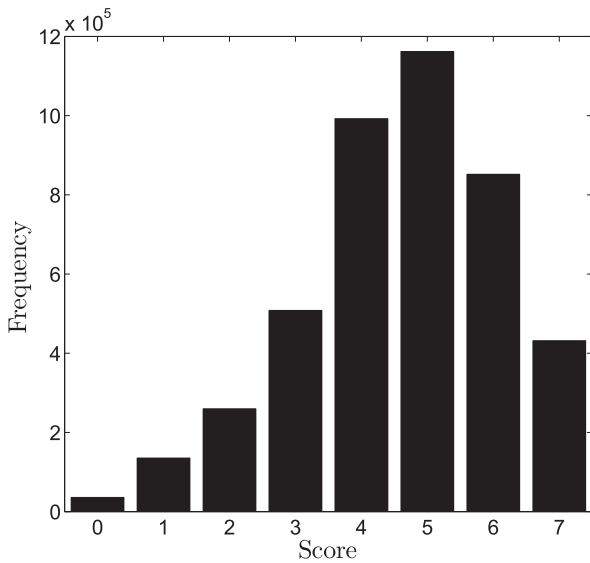


図 1: レビュー評点の分布

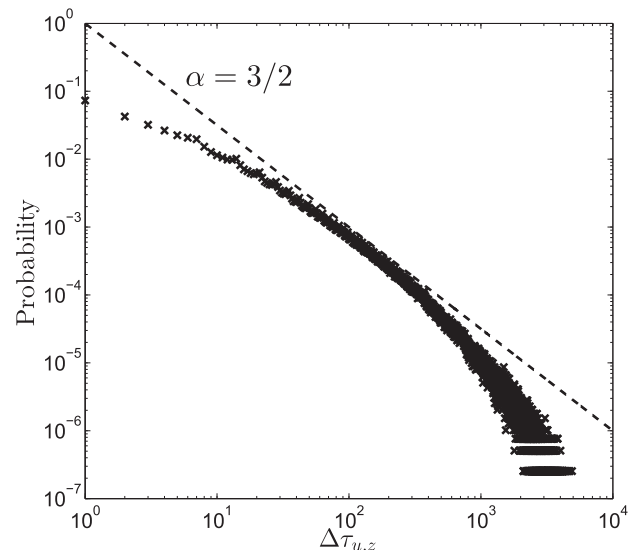


図 2: ユーザのレビュー投稿間隔分布

までの時間間隔の確率分布は  $f(x) = x^{-\alpha}$  ( $\alpha = 3/2$ ) の冪乗則に近似することが Oliveira と Barabasi によって既に報告されている [4] ため、検証にはユーザのレビュー投稿間隔集合を用いる。

レビューデータに含まれるユーザ  $u$  とユーザ集合  $U$  を

$$u \in U = \{u_1, \dots, u_M\} \quad (7)$$

と定義する。このとき、ユーザ  $u$  がレビューを投稿した時刻を  $\tau_u$  とし、時刻の新しいものから  $\tau_{u,1}, \tau_{u,2}, \dots$  としていくと、ユーザ  $u$  のレビュー時刻集合  $I(u)$  は

$$I(u) = \{\tau_{u,1}, \dots, \tau_{u,|I(u)|}\} \quad (8)$$

となり、ユーザ  $u$  のレビュー投稿間隔集合  $\Delta I(u)$  は、

$$\Delta I(u) = \{\Delta\tau_{u,z} = \tau_{u,z} - \tau_{u,z-1} : z = 2, \dots, |I(u)|\} \quad (9)$$

となる。よって、ユーザのレビュー投稿間隔集合  $I$  は、

$$I = \bigcup_{u \in U} \Delta I(u) \quad (10)$$

で書き表すことができる。

図 2 は、データセットから求めたレビュー投稿間隔集合  $I$  の  $\Delta\tau_{u,z}$  を日数に変換したときの確率分布であり、破線は  $f(\Delta\tau_{u,z}) = \Delta\tau_{u,z}^{-\alpha}$  ( $\alpha = 3/2$ ) を示す。図より、ユーザの投稿間隔の確率分布は、 $\alpha = 3/2$  の冪乗則に近い分布となっているため、このデータセットのレビューには、人間の行動特性が少なからず反映されていると言える。

#### 4.2 実験結果

データセットのアイテム毎で得られた分割点集合  $C_K$  に関する結果を示す。ここで、 $C_K$  はシンプル法で求めた分割点集合をムーブ法で改善したものであり、シンプル法の  $\chi^2$  の棄却点は  $p = 0.005$ 、自由度 7 の 20.2777 である。

図 3, 4, 5 に、 $K$  の度数分布、 $k$  と  $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  のプロット、 $K$  と被レビュー数  $N$  のプロットをそれぞれ示す。また、表 1 には  $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  の上位 10 アイテム

を示す。図 3, 4 より、 $K$  の度数は冪乗則に従っており、 $k$  と  $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  の関係も似たような性質を持っていることが分かる。また、図 5 より、分割点数  $K$  はアイテムの被レビュー数  $N$  に大きく依存していることが分かる。なお、この実験で得られた分割点の総数は 8119 である。

さらに、図 6 と 7 に、 $K$  ごとの解の改善値平均と、 $K$  ごとの計算量の増加比率平均を示す。図 6 の縦軸は、ムーブ法のプロセスにおいて改善された目的関数値の平均を表しており、また、図 7 の縦軸は、ムーブ法によって増加した計算量の比率（全体の計算量/シンプル法のみ計算量）の平均を表している。図 6 より、 $K$  の増加と共に改善値が増加することは明確であり、また、図 7 より、全体の計算量はシンプル法の高々 3 倍程度にしかならないことが分かる。

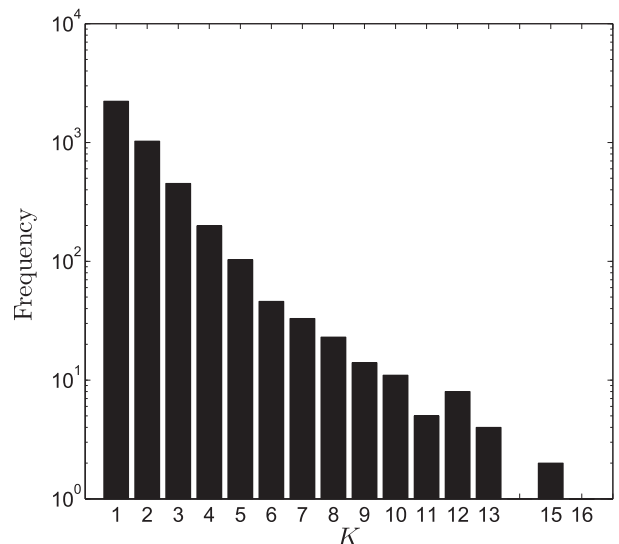


図 3:  $K$  の度数分布

#### 4.3 分割点分析

ここでは、 $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  の上位アイテムに対し、詳細な分析を行う。表 1 から分かるように、上位のアイテ

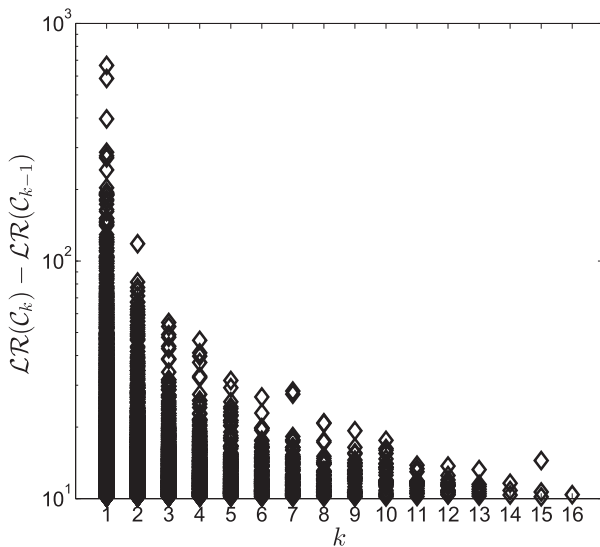


図 4:  $k$  と  $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  のプロット

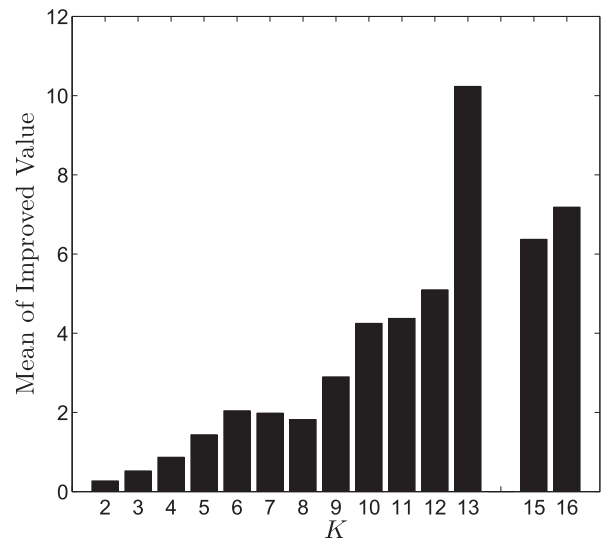


図 6:  $K$  ごとのムーブ法による解の改善値平均

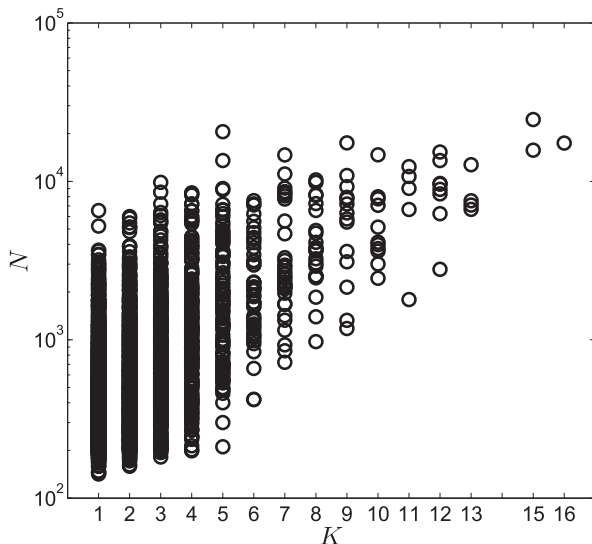


図 5:  $K$  と  $N$  のプロット

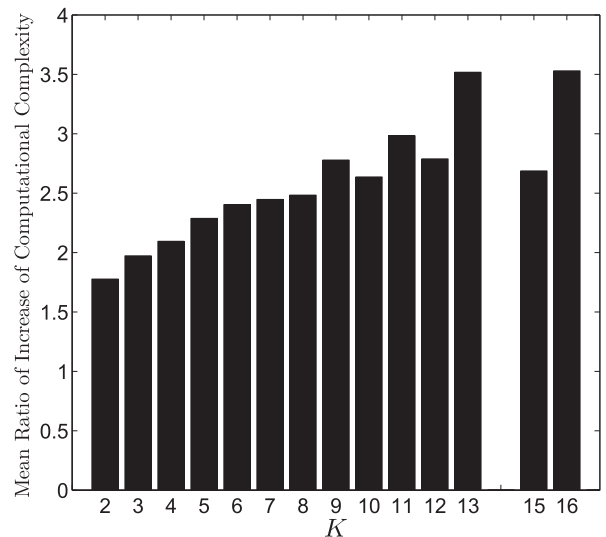


図 7:  $K$  ごとのムーブ法による計算量の増加比率平均

ムの被レビュー数  $N$  は非常に多く、図示した際の視認性が非常に悪いので、今回は被レビュー数  $N$  が 500 以下のアイテムを紹介する。以下に、 $N \leq 500$  という条件において、 $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  が最も上位であった 2 アイテムの分割点分析について述べる。

1 つ目はカットコットン(無印良品)というアイテムで、全分割点中  $T_1$  が 78 位、 $T_2$  が 5626 位である。図 8 は評点の推移と  $T_k$  を示したものであり、グレーの実線は評点の推移を、黒の実線は評点の移動平均 ( $s_n + s_{n-1} + \dots + s_{n-6}/7$ ) を、破線は  $T_k$  をそれぞれ表す。 $T_1$  以降は高評価で安定しているので、実際に  $T_1$  前後 10 レビュー程度を確認すると、アイテムそのものの改良が大きく影響していることが分かった。より具体的には、「毛羽立ちが抑えられている」や「肌触りが良くなった」という意見が多く見られ、枚数を増やしたのに価格を据え置きにしていることについても高い評価を得ていた。この分割点には、情報に敏感なユーザ (Early

Adopters) が大きく関わっていることが分かる。次いで  $T_2$  前後 10 レビュー程度を確認すると、初期のレビュー評価が低すぎると感じたユーザによる対抗レビューが  $T_2$  以降多く書かれており、評価の回復傾向が見られた。

2 つ目は魔法の泡立てネット(モンクレール)というアイテムで、全分割点中  $T_1$  が 81 位、 $T_2$  が 1410 位、 $T_3$  が 3132 位である。先程と同様、図 9 に評点の推移、評点の移動平均、 $T_k$  を示す。このアイテムの場合、 $T_1$  から評価の低迷が見られるので、実際に  $T_1$  前後 10 レビュー程度を確認すると、他社が頑丈なものを低価格で販売し始めたことが影響していることが分かった。さらにレビューを過去に遡って見ていくと、発売当初は泡立てネットそのものが市場に存在していなかったため、商品の目新しさによって高評価のレビューが集まったことも分かった。類似商品が出現し始めた  $T_1$  以降は、それらと比較して低評価としているレビューが多く、発売当初とのギャップが大きい。先程と同様、この

表 1:  $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  上位 10 アイテム

No.	Item Name	$k$	Value	$T_k$	$N$
1	大島椿 (ツバキ油)	1	665.35	5095	24552
2	オートマテック ライナー	1	586.70	6841	7730
3	シャドーカスタマイズ	1	396.15	2379	11127
4	NU ソワン オレオリラックス	1	286.92	4905	9853
5	みつばちマーチ	1	277.13	5656	12380
6	シアトリカルパウダー	1	271.77	2117	9590
7	エクスプレスケア トータルクリーン	1	241.82	197	2196
8	ビューティー コスメマスカラ	1	203.34	3446	8382
9	資生堂眉墨鉛筆	1	194.65	4460	7056
10	ロングウェア ジェルアイライナー	1	192.04	2380	7759

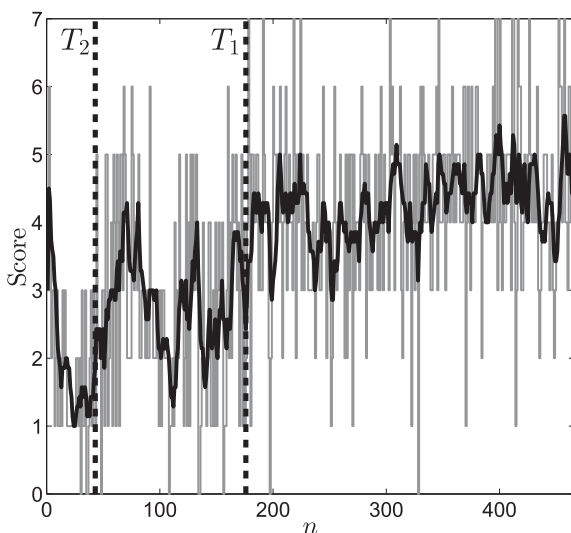


図 8: カット Cotton の評点推移と  $T_k$

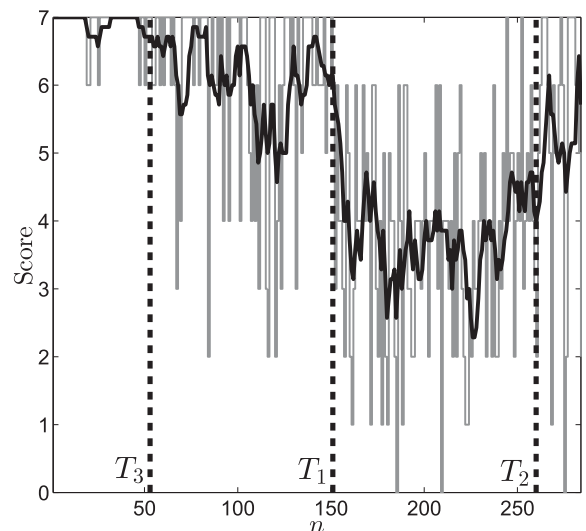


図 9: 魔法の泡立てネットの評点推移と  $T_k$

分割点も、情報の持ち込みが早いユーザ (Early Adopters) によって引き起こされたものと見られる。一方、 $T_2$  前後 10 レビュー程度とそれ以降では、「長期間使用した」とレビューに明記しているユーザの高評価レビューが出現し始め、評価にバラつきが生じている。 $T_3$  から  $T_1$  の間は、辛口のレビュアーによる厳しい批評も登場するが、レビュー傾向は序盤と同様、高評価寄りであった。

一方、 $N$  が非常に多いアイテムは、図 5 から分かるように、分割点数  $K$  も比較的多くなるため、自然と分割統治法のように問題が細分化する。つまり、 $N$  が非常に多い場合でも、 $N$  が少ないアイテムと同様な分析が可能となる。 $\mathcal{LR}(C_k) - \mathcal{LR}(C_{k-1})$  上位のアイテムの分割点分析では、今回紹介したようなイベント以外にも、類似品との誤認識、低年齢層ユーザによる一斉批判、影響力のあるデマによる混乱、やらせ・サクラの類等の多種多様なイベントが見られた。

ここで、提案法による分割結果は、レビューの投稿時刻や投稿間隔といった時間的な密度を一切考慮していないことに注意されたい。言い換えれば、これらの分割結果は、アイテムがいつ注目されたか等の「話題性」を一切考慮して

いないということであり、逆に考えれば、「話題性」に左右されていないということである。

### 5. 計算量と解品質

計算量と解品質の関係をさらに追求するべく、二種類の解法を比較する。

#### 5.1 逐次法

4.2 項の実験時と同様の解法を逐次法とする。すなわち、その手順は以下となる。

- R1. シンプル法で  $C_K$  を求める。
- R2. ムーブ法で  $C_K$  を改善する。

#### 5.2 相互法

$C_k$  が更新される度に解の改善を図る解法を相互法とする。すなわち、その手順は以下となる。

- R1. シンプル法の S5 の処理開始時に  $k \geq 2$  なら、 $C_k$  を  $C_K$  として出力。
- R2. ムーブ法で  $C_K$  を改善し  $C_K$  を  $C_k$  として出力。
- R2. シンプル法の S5 から処理を再開し、R1 へ戻る。

#### 5.3 解法の比較

以下に、4.1 項の @cosme データセットにおける解法の比較結果を述べる。図 10, 11, 12 はそれぞれ、 $K$  の度数分布、 $K$  ごとの解の改善値平均、 $K$  ごとの計算量の増加比率平均についての比較を示したものであり、黒棒が逐次法の結果、白棒が相互法の結果である。

図 10 より、相互法で得られる分割点数の期待値は、逐次法よりも低いことがわかる。分割点の総数で見ても、逐次法が 8119 だったのに対し、相互法は 7930 であった。また、図 11 より、相互法は逐次法に比べて解の改善精度が高い傾向にあるので、相互法は高い解品質が期待できる。一方、図 12 より、相互法の計算量は約  $K$  倍に膨れ上がっているため、逐次法の方が、さらに大規模なデータでも高速に解けることが期待できる。

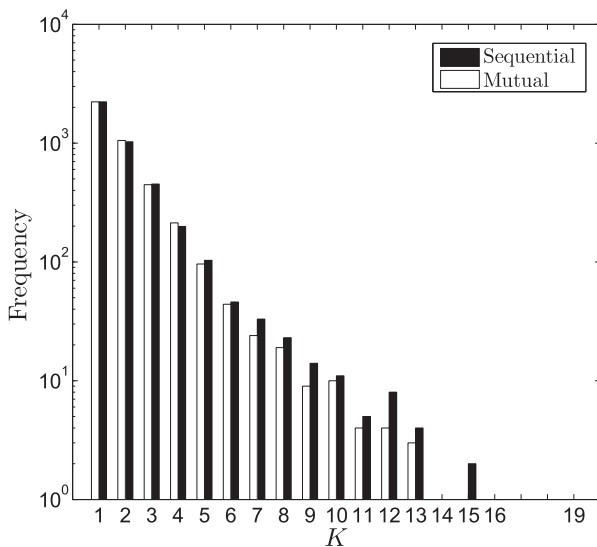


図 10:  $K$  の度数分布の比較

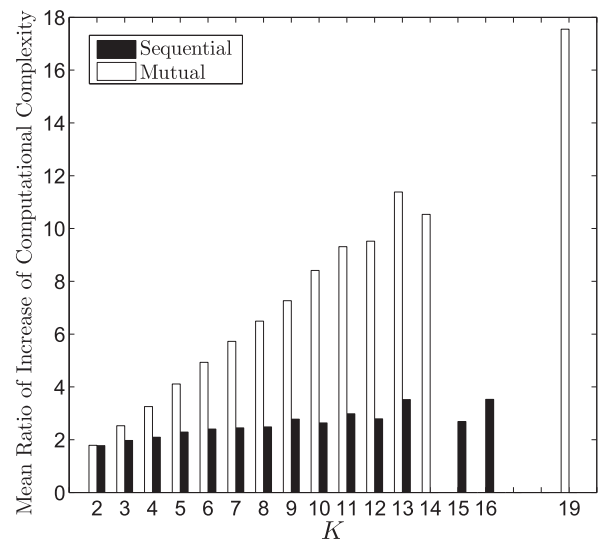


図 12:  $K$  ごとの計算量の増加比率平均の比較

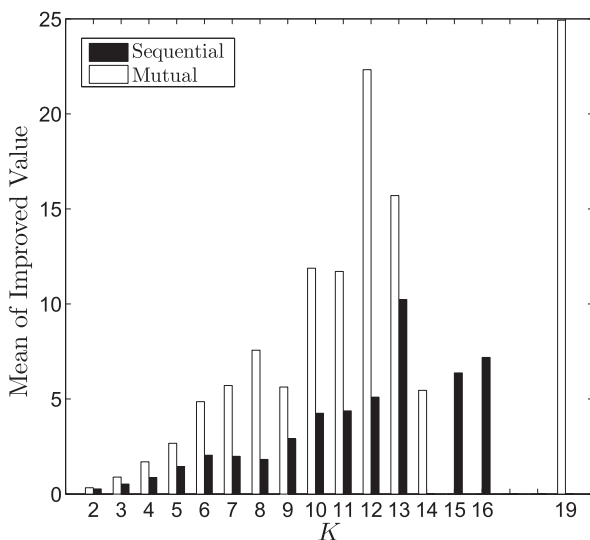


図 11:  $K$  ごとの解の改善値平均の比較

## 6. おわりに

評点時系列データからのイベント検出を目的として、区間分割問題とその解法を提案し、大規模レビューデータを用いて実験を行った。問題設定では、尤度比検定の枠組みによって区間分割問題を定式化し、また、解法の提案では、高速且つ高い解品質が期待できるアルゴリズムを考案した。実験では、提案法がレビューの分析支援として十分有用であることを示した。さらに、今回提案した解法における解品質と計算量の関係についても深く追求した。今後は、更なる大規模データや、他の時系列データへの適応を試みる予定である。

## 【謝辞】

本研究は、豊田中央研究所との共同研究、及び、科学研究費補助基金基盤研究(C)(No.25330635)の支援を受けて行ったものである。

## 【文献】

- [1] Salganik, M. J., Dodds, P. S., and Watts, D. J., "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, vol.311, pp.854–856, 2006.
- [2] Swan, R., and Allan, J., "Automatic Generation of Overview Timelines," *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.49–56, SIGIR '00, 2000.
- [3] Kleinberg, J., "Bursty and Hierarchical Structure in Streams," *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol.7, pp.91–101, 2002.
- [4] Oliveira, J. G., and Barabasi, A. L., "Human dynamics: Darwin and Einstein correspondence patterns," *Nature*, vol.437, pp.1251, 2005.

## 山岸 祐己 Yuki YAMAGISHI

静岡県立大学大学院経営情報イノベーション研究科修士課程在学中。データマイニングの研究に従事。人工知能学会、情報処理学会、日本データベース学会各学生会員。

## 斉藤 和巳 Kazumi SAITO

静岡県立大学経営情報学部教授。1985年慶応義塾大学理工学部数理工学科卒業、同年日本電信電話(株)入社。1998年東京大学博士(工学)。複雑ネットワーク、機械学習の研究に従事。情報処理学会、電子情報通信学会、人工知能学会、日本神経回路学会、日本応用数理学会、日本行動計量学会、日本データベース学会各会員。

## 武藤 伸明 Nobuaki MUTOH

静岡県立大学経営情報情報学部准教授。1995年横浜国立大学大学院工学研究科博士課程後期修了。博士(工学)。離散幾何学、組合せ最適化の研究に従事。電子情報通信学会他会員。