

評価値の時系列変化に着目したユーザレビューの分析

高橋 毅[†] 天笠 俊之^{††,†††} 北川 博之^{††}

[†] 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 筑波大学システム情報系 〒 305-8573 茨城県つくば市天王台 1-1-1

^{†††} 宇宙航空研究開発機構 宇宙科学研究所 〒 252-5210 神奈川県相模原市中央区由野台 3-1-1

E-mail: [†]s1120717@u.tsukuba.ac.jp, ^{††}{amagasa,kitagawa}@cs.tsukuba.ac.jp

あらまし 旅行先の宿泊施設を選ぶ際に宿泊施設レビューサイトを利用するユーザ数が増えつつある。宿泊施設レビューには宿泊者が宿泊施設に対して与える評価値の他に、該評価を与えた理由が記述されている事がある。理由を発見して利用者に提示する事は、業務改善など有益な利用に繋がると考えられる。また理由は常に同じとは限らず、設備の故障や施設の改築など不定期に発生するイベントが要因として挙げられるケースが存在する。本研究では、レビューが宿泊施設に与える評価値が時系列変化する点に着目し、変化のあった時間前後に投稿されたレビューを分析する事によって、該宿泊施設の評価に影響を与えた要因を示すレビューを検出する。

キーワード データマイニング, レビュー, 時系列

Tsuyoshi TAKAHASHI[†], Toshiyuki AMAGASA^{††,†††}, and Hiroyuki KITAGAWA^{††}

[†] Graduate School of Systems and Information Engineering, University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, Japan

^{††} Division of Information Engineering, Faculty of Engineering Information and Systems, University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, Japan

^{†††} Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency

3-1-1 Yoshinodai, Chuo-ku, Sagami-hara-shi, Kanagawa 252-5210, Japan

E-mail: [†]s1120717@u.tsukuba.ac.jp, ^{††}{amagasa,kitagawa}@cs.tsukuba.ac.jp

1. はじめに

近年、レビューサイトの重要性が商業的にも学術的にも注目されている。例えば、宿泊施設のレビューサイトを例にとると、そこには各施設に対する多数の利用者からの評価とレビュー文が蓄積されている。別の利用者が宿泊施設を探す際は、施設自身が提供する情報だけでなく、レビュー結果を参照する事により、利用者の視点に立った評価を参考にする事が出来、より満足度の高い施設選択を行えるようになる。

一方、レビューデータは利用者だけではなく、宿泊施設にとっても有益な情報になる。例えば、評価値から顧客の満足、不満な点を洗い出し、業務改善に活かす事が出来る。このため、膨大なレビューデータを分析し、そこから有用な情報を抽出することが重要となる。そのような分析方法には多様なアプローチが考えられるが、一つの考えとしてレビュー結果の時間的な変化に着目して、評価値が上がった（下がった）あるいは記述されたレビューの内容が変化した時刻に着目し、その原因を探る

事が考えられる。

本研究では、レビュー結果の時間変化に着目し、その原因を分析する手法を提案する。このため、まずレビューが宿泊施設に与える評価値に着目してレビューの選別を行う。レビューによっては、有効な評価をせず決まった値のみを入力する場合が考えられる。そのような評価値を与えるレビューによるレビューはノイズとしてあらかじめ取り除くことが望ましい。この目的のために評価値のエントロピーに着目する。次に、有効な評価結果に基づいて評価値の時系列的変化を観察し、変化のあった期間を発見する。変化のあった期間にレビューが述べているトピックに注目し、トピックと関連するレビューを検出する事で、要因を分析する。

本論文の構成は以下の通りである。まず2節で関連研究について述べる。次に、3節でレビューデータの詳細について述べ、4節で提案手法について述べる。5,6節では得られた結果について考察を与える。

2. 関連研究

レビューを対象にした研究は盛んに行われている．関連研究として，レビュー文を要約して意見を抽出する研究，レビューデータにトピックモデルを適用した研究，レビュー結果の時系列変化に着目した研究を挙げる．

膨大なレビューデータを要約し，詳細な意見を抽出するための研究が自然言語技術の分野で盛んに行われている [2]．辞書を作成して意見を抽出する研究として [3]～[5] 等が挙げられる．小林らは意見を {評価対象，属性，評価値} の三つ組の評価表現形式に構造化し，レビュー文を係り受け解析して評価表現を抽出する事で意見を要約した [3]．辻井らは小林らの手法を用いて特定の地域内の宿泊施設に対するレビューを分析し，旅行先に応じてレビューが重要視する情報が異なる事を示した [4]．藤井らは表現が好評と不評のどちらを表すかを示す評価表現辞書を作成し，評価表現を抽出．各評価表現を用いてホテルの評価値を予測した．更に予測したホテルの評価値と評価表現を二次元マッピングする事で，ホテルに与えられた評価の要因を効率的に分析する手法を提案した [5]．しかし藤井らの研究では与えられる評価値が時系列変化する事を考慮していない．

レビューデータにトピックモデルを適用した研究として [6], [8] 等が挙げられる．北島らは，レビュー文を {対象-評価} で表される事象の集合として捉え，レビューデータに LDA を適用した．レビュー文を単語の集合として捉えた場合に比べて，文書の内容をより明確に捉えたトピック推定を行える事が示された [6]．Titov らは，レビューデータに LDA を適用して得られたトピックを評価項目に分類し，言及している評価項目毎にレビューを分類した [8]．我々は各レビューが何を表すかを知るためにトピックモデルを利用する．レビューが表す内容は評価項目だけに限らないため，レビューを評価項目毎に分類するのは好ましくない．

レビュー結果の時系列変化に着目した研究として [9] が挙げられる．山岸らは，評価対象に不自然な評価が与えられている期間を発見する手法を提案し，期間内に投稿されたレビューを分析した．分析によって期間内に信頼度の低いユーザが集中して投稿を行っていた事を示し，手法の有効性を評価した [9]．山岸らの手法では評価対象の評価に影響を与えた要因として，レビューによる投稿に焦点が当てられている．我々が目的とする要因はレビューの振る舞いではなく，レビュー対象に関係して発生した事象とする．

本研究では，有用なレビューを投稿するレビューアを，登録した評価値のエントロピーで評価し，ある評価値（全て 3，全て 5 など）だけを登録するようなレビューを排除する．また，評価値の時系列的な変化に着目して，その原因をレビュー文から分析する．

3. レビューデータ

本研究では楽天株式会社から提供された楽天トラベルのレビューデータ [1] を用いた．レビューデータには，レビュー対象となる宿泊施設の名称を示す「施設名」，レビューアが自由

記述した「レビュー文」，レビューが投稿された「投稿時間」マスクされた「利用者名」，レビューの内容を簡潔にまとめた「施設回答文」，レビュー ID を示す「投稿番号」等の他に，評価項目として立地，部屋，食事，風呂，サービス，設備・アメニティ，総合が与えられている．レビューによって 0-5 の 6 段階の評価値が各評価項目に与えられる．

表 1 例，各期間におけるトピック寄与率

項目	例
施設名	ホテル A
レビュー文	「久々に利用しました…」
利用者名	user00001
投稿番号	7100000
投稿時間	2010 年 8 月 10 日 8:10:00
施設回答文	「…ご指摘の件につきましては…」
立地評価値	4
部屋評価値	4
食事評価値	0
風呂評価値	3
サービス評価値	5
設備・アメニティ評価値	2
総合評価値	3

4. 提案手法

本研究では実世界でのイベントを検出するために，評価値が時間変化した点に着目する．また，その原因を特定するため，変化点の前後におけるレビュー内容の変化にも着目する．

4.1 レビューアの選別

まず，レビューにおける数値評価の変化に着目したイベント検出を行う．このため，6 段階評価における数値の分布を調査した．この結果，レビューアが与える評価値は 4 または 5 に偏る傾向が強いことが分かった．これは楽天トラベルだけではなく，他のレビューサイトにも共通している性質だということが知られている [4], [5]．このような結果が得られる原因の一つとして，偏った評価を与えるレビューアが存在が挙げられる．特に不満がない場合，少なくない割合のレビューアが最高得点かそれに近い値を一貫して与えることがある．また，レビューア間の慣例として，そのようなスコアの与え方が推奨されていることもある．

このような偏ったスコアを与えるレビューアからの評価は，正当な評価であるとは考えられないため，イベント検出処理の前に除外したい．このため本研究では，レビューアの与えるスコアからエントロピーを計算し，一定の閾値を基準にレビューアのフィルタリングを行う．

レビューアが各評価項目に与える評価値の確率を用いてエントロピーを算出する．レビューアがレビューした回数を t ，評価項目 k に対して評価値 i を与える回数を t_{ki} とすると評価項目 k に対して評価値 i を与える確率 $p_{ki} = t_{ki}/t$ となる．よって，レビューアの評価項目 k におけるエントロピー E_k は式 1 となる．

$$E_k = \sum_{i=0}^5 (-p_{ki} \times \log_2 p_{ki}) \quad (1)$$

レビューが持つエントロピー E_{sum} を E_k の総和として式 2 に定義する.

$$E_{\text{sum}} = \sum_{k=0}^6 E_k \quad (2)$$

式 1, 式 2 によって得られた E_{sum} をレビューのエントロピーとする. 得られたエントロピーを基準に, レビューの選別を行う. 本研究では, E_{sum} が 0 でないレビューを処理の対象とするが, 閾値はデータセットなどに応じて適切に設定する必要がある.

4.2 イベント発生期間の推定

本節ではイベントが発生した時期の推定方法を述べる. 提案手法では, 評価期間毎にイベントが発生したかどうかを推定するため, 事前に各宿泊施設において評価期間を設定する必要がある. まず評価期間の決め方について述べ, 次にイベントの推定方法について説明する.

4.2.1 期間の長さ

イベントの検出を発見する前にレビューが投稿される間隔を観察した. 期間の長さに着目して投稿件数を観察した結果, 短期間に多くレビューされる宿泊施設と, 1ヶ月間に 2,3 件しかレビューされない宿泊施設が確認された. またレビュー件数に着目して評価期間を観察した結果, 一定のレビュー件数が投稿されるまでに 2ヶ月以上掛かる宿泊施設を確認した. これは, 宿泊施設によって立地や利便性が異なるため, 宿泊者の数に差が出た事が原因と考えられる.

宿泊施設毎に, レビュー件数と評価期間の長さに応じて, 評価期間を設定する必要がある.

これを解決するために, 評価期間 T の長さ, 直前の評価期間と重複する期間の長さ, 期間内のレビュー件数について条件を設ける. T の長さの条件を以下の通りにする.

- 評価期間の長さ: 2 日以上
- 重複する期間の長さ: 直前の評価期間の長さに応じる

イベントが評価期間をまたぐように発生する場合を考慮し, 重複期間を設ける. 重複期間は期間内に投稿されるレビューを考慮して決定する. 直前の評価期間 $T-1$ におけるレビューの集合を R^{T-1} , 評価期間 T におけるレビューの集合を R^T として, 以下の流れによって決定する.

- (1) $T-1$ の後半を重複期間 $\frac{T-1}{2}$ とする.
- (2) 包含関係 $R^{T-1} \supset R^T$ を満たさなくなるまで, 以下の流れを繰り返す.
 - (a) $\frac{T-1}{2}$ を含めた T における R^T を得る.
 - (b) 包含関係を満たす場合, $\frac{T-1}{2}$ の後半部分を新たに重複期間とする.

評価期間の条件により, 一日以上の重複期間が含まれる. また T に投稿されるレビュー件数 r に対して, $r_{\min} \leq r < r_{\max}$ を条件として定める. ただし以下の場合には例外として期間, レビュー件数に制約を設けない.

- 2 日間でレビュー件数が $r_{\max}$ 以上投稿される.
- 1 ヶ月で経っても $r_{\min} \leq r$ を満たさない.

以上の条件を基に, 期間の長さとレビュー件数に応じて T を以下の三通りに分類する.

- 2 日間で $r_{\max}$ 以上のレビューが投稿される期間.
- 期間の長さが 2 日以上 1 ヶ月以内で, レビュー件数が $r_{\min}$ 以上 $r_{\max}$ 未満を満たし, かつ最長となる期間.
- 1 ヶ月経っても, レビュー件数が $r_{\min}$ 以下となる期間.

$r_{\min}$, $r_{\max}$ はデータセットに応じて適切に与える必要がある. 本論文では, $r_{\min} = 5$, $r_{\max} = 10$ とした.

4.2.2 イベント検出

実世界で何らかのイベントが発生した場合, それがレビューの評価結果に影響を与え, 評価値に影響を受けることが考えられる. 評価値の変動に着目すれば, 実世界で何らかのイベントの発生を検出することが期待される. 評価値によるイベントの検出を行うには, 単純にはある施設に関するレビューの評価値を時期ごとにグループ化し, その平均値の変動を利用することが考えられる. しかし, 実際には平均値の変動は大きく, それだけを使ってイベントを検出することは難しい.

本研究では, 評価値平均の変動ではなく, ばらつきを検出する. 何らかのイベントによって一部のレビューが影響を受けたとすると, それは評価値のばらつきに反映されると考えられる. 本研究では各宿泊施設において評価期間を設定し, 各期間における評価値の分散を計算する. 宿泊施設において分散が高い期間をイベントが発生した期間として推定する.

4.3 変動要因の推定手法

4.3.1 手法の概要

評価値の分散の変動に着目することで, イベントを検出することができた. 次は, そのイベントが発生した要因を分析する. これにはレビュー文を利用する. レビュー文を利用した原因分析には, おおまかに次の二つのアプローチが考えられる.

- (1) イベントが検出された時期に投稿されたレビューのうち, 他と比べて評価値が低いレビューのレビュー文を参照する.
- (2) 各レビューがもつトピックを捉え, 各期間におけるトピックの傾向を捉える. 他の期間よりも多く記述されているトピックを変動要因として捉え, 関連するレビュー文を参照する.

4.2 節で述べたように, 投稿される評価値のほとんどが高い値を示すため, 変動要因は低い評価値の投稿である可能性が高い. (1) によって, 変動要因を示すレビューを検出できると考えられる.

不定期に発生するイベントの影響で評価値に影響が出た場合, レビューがイベントをレビューのトピックとして記述する可能性が高い. また, 変動要因を示す内容は低い評価値を与えるレビューだけではなく, 高い評価値を与えるレビューにも含まれている可能性がある. (2) では, トピックに注目してレビューを参照する事で, 低い評価値に限定せずに変動要因を探し出せると期待される.

以下では, (2) に基づいて変動要因に関連するレビューを検出する方法について述べる.

4.3.2 トピックに着目した検出手法

提案手法の流れを説明する。

- (1) 4.1 節で説明した方法でレビューを選択する。
- (2) 選択されたレビューによって投稿されたレビューデータに対して LDA を適用し、各レビューのトピック分布を求める。
- (3) 評価期間毎に投稿されたレビューに対するトピック分布を合計し、評価期間毎に各トピックの特徴量を得る。
- (4) 各トピックについて、イベントが検出された期間における特徴量と前後の期間における特徴量を比較し、スコアを算出する。
- (5) 各トピックのスコアと、該当期間に投稿された各レビューにおけるトピック寄与率を比較し、イベントに関連するレビューを検出する。

スコアについて説明する。イベントが発生した場合、イベントを示すレビューが投稿される可能性がある。レビューの投稿によって、イベントに関連するトピックの特徴量が前後の期間よりも大きくなると予想される。このようなトピックを発見するために、該当期間では前後の期間に比べてトピックがどれだけ述べられているのかを示すスコアを求める。具体的には、イベントの発生した期間における各トピックの特徴量と前後の期間における各トピックの特徴量を比較してトピックのスコアを求める。

イベントが発生した期間 T におけるトピック t の特徴量を p_t^T 、直前の期間 $T-1$ における t の特徴量を p_t^{T-1} 、直後の期間 $T+1$ における t の特徴量を p_t^{T+1} として、 T における t のスコア $score_t^T$ を式 3 によって求める。

$$score_t^T = \frac{p_t^T}{p_t^{T-1}} + \frac{p_t^T}{p_t^{T+1}} \quad (3)$$

本研究では、イベントが発生した期間におけるトピックのスコアと各レビューのトピック寄与率を比較し、変動要因を示すレビューを検出する。実験では順位相関、Jensen-Shanon 情報量によって変動要因を示すレビューを検出し、比較する。

4.3.3 順位相関による検出

イベントが発生した期間において、以下の二つの順位の順位相関を求める。

- スコアによるトピック t の順位 $rank_t^{score}$
 - レビュー r における寄与率によるトピック t の順位 $rank_t^r$
- n を全トピック数として、式 4 によって順位相関係数を求める。

$$rank_correlation = 1 - \frac{6 \sum (rank_t^{score} - rank_t^r)^2}{n(n^2 - 1)} \quad (4)$$

順位相関係数が高いレビューを、変動要因と関連が強いレビューとして検出する。

4.3.4 Jensen-Shanon 情報量による検出

Jensen-Shannon 情報量は二つの確率分布の差異を表す指標の一つである。類似度を判定する際に用いられ、[3], [7] では最

も精度の高い結果が得られている。Jensen-Shannon 情報量は、同じく確率分布間の差異を示す指標である Kullback-Leibler 情報量によって求められる。二つの確率分布 P, Q において、Kullback-Leibler 情報量 D_{KL} は式 5 で表される。

$$D_{KL}(P||Q) = \sum_i P(t_i) \log \frac{P(t_i)}{Q(t_i)} \quad (5)$$

Kullback-Leibler 情報量は非対称的に確率分布間の差異を表すが、Jensen-Shannon 情報量是对称的に確率分布間の差異を表す。評価期間 T において、スコアの分布を d_{score} 、該当期間に投稿されたレビュー r におけるトピック分布を d_r としたとき、Jensen-Shannon 情報量は D_{KL} を用いて式 6 で表される。

$$D_{JS}(d_r||d_{score}) = \frac{1}{2} D_{KL}(d_r||\frac{d_r + d_{score}}{2}) + \frac{1}{2} D_{KL}(d_{score}||\frac{d_r + d_{score}}{2}) \quad (6)$$

順位相関ではスコアによる順位を全て用いてレビューの検出を行った。しかしトピックの中には特徴量が殆ど変化しないトピックが存在する。このようなトピックは変動要因と関係しない可能性が高い。事前に変動要因を示すトピックを選別するために、スコアが高いトピックに注目する。変動要因を示すトピックのみを用いる事で、イベントを示すレビューを適切に検出できる可能性がある。

5. 評価実験

本節では分析結果について説明する。実験では 3 節で述べた楽天トラベルのレビューデータのうち、2009 年 4 月 1 日以降に投稿されたレビュー 2028274 件を実験に利用した。本手法では LDA を適用するため、各レビュー文を事前に形態素解析する必要がある。実験ではレビュー文を Mecab [12] によって形態素解析し、単体で意味を判別できる品詞として以下の品詞のみを抽出した。

- 名詞
 - サ変接続
 - ナイ形容詞語幹
 - 形容動詞語幹
 - 一般名詞
 - 固有名詞
- 未知語

品詞の判別には IPA 品詞体系を利用した。また、本実験では Mallet [13] によって LDA を適用した。

5.1 レビューの選別

レビューの選別結果について述べる。データセットから、 E_{sum} が 0 より大きいレビュー 220169 名を検出した。 E_{sum} が 0 より大きいレビューと 0 のレビューが与えた評価値の分布を表 2 に示す。表 2 より、 E_{sum} が 0 より大きいレビューは評価値 2 以上を与える傾向があり、 E_{sum} が 0 であるレビューは評価値 1 以下を与える傾向がある事が分かった。特に、評価値 0 を与えているレビュー数の多くが E_{sum} が 0 であるレビューによる

レビューである事が分かった。

本手法においてエントロピーに着目してレビューを選別する理由は、高い評価値を与えるレビューを除く事ではなく、偏った評価を与えるレビューによるレビューを除くためである。そのため、 E_{sum} が 0 より大きいレビュー 220169 名が投稿した 1096726 件のレビューを利用する。なお、 E_{sum} が 0 であるレビューには、一度しかレビューをしていないレビューも含まれる。

表 2 レビューが与えた評価値の分布

評価値 (総合)	全レビュー	$E_{\text{sum}} > 0$	$E_{\text{sum}} = 0$
0	13494	187	13307
1	28915	12987	15928
2	62109	31146	30963
3	245194	139828	105366
4	933853	522887	410966
5	744709	389692	355017

5.2 変動要因の分析

変動要因の分析の結果について述べる。まず最適なトピック数、各トピックのキーワードについて述べる、次に各評価期間における各トピックの特徴量を観察した結果について述べる。レビューの検出精度について述べた後、スコアと検出されたレビューの分析について述べる。

5.2.1 抽出されたトピック

LDA ではトピック数を事前に与える必要がある。トピック数を決める方法の一つとして、評価項目をトピックとして利用する事が考えられる。しかしレビューが各評価項目に関してのみレビュー文を書くとは限らないため、最適なトピック数を推定する必要がある。

最適なトピック数の決め方は北島ら [6] の論文を参考にする。北島らは、文書検索精度が最も高いトピック数では各文書のトピックを適切に捉えていると考えた。本手法では LDA によって取得したトピックをレビューが持つ話題として捉えるため、各レビューが持つトピックを適切に捉えているトピック数が最適なトピック数と考えられる。そのため、本論文でも文書検索精度によって最適なトピック数を決定する。

実験では以下のように実験環境を設定した。

- クエリ：単語「部屋」
- 正解データ：施設回答文に「部屋に関して」と記述されているレビュー文から 1000 件
- テストデータ：施設回答文に単語「部屋」を含まないレビュー文から 1000 件
- 訓練データ：正解データ、テストデータに含まれないレビュー文 10000 件

クエリを「部屋」とした理由は、多くのレビューにおいて部屋に関するコメントが含まれるためである。施設回答文にはレビューの話題を要約した内容が記述されているため、正解データ、テストデータの基準として利用した。実験ではトピック数を 100,125,150,175,200 と変えて調査を行う。トピック数毎に

10 回ずつ以下を繰り返す。

- (1) 訓練データに LDA を適用し、トピック抽出。
- (2) クエリ、正解データ、テストデータにおけるトピック分布を求める。
- (3) 正解データ、テストデータにおけるトピック分布とクエリにおけるトピック分布の Jensen-Shannon 情報量を求め、低い順に並べる。
- (4) 11 点平均適合率を求める。

算出した 11 点平均適合率の結果をトピック数毎に平均し、該トピック数における文書検索精度とした。結果を表 3 に示す。

表 3 各トピック数における 11 点平均適合率の平均

トピック数	11 点平均適合率の平均
100	0.515134
125	0.515067
150	0.520775
175	0.516564
200	0.513308

結果より、トピック数 150 のときの適合率が最も高いことから、最適なトピック数を 150 とした。なお、hdp-lda [10] によって、最適なトピック数を大まかに推測する事も可能である。

次に得られた各トピックのキーワードを嘉村ら [11] の提案する PDTD-2 を利用して抽出する。本手法では、トピックが宿泊施設の評価に影響を与える要因として捉える。トピックを客観的に判断できる状態になっている事が望ましいため、トピックのキーワードを決めた。作成された各トピックのキーワードと生成確率の高い上位 5 件の単語を表 4 に示す。なお、紙面の都合により 50 件のトピックについてのみ示す。

以下では宿泊施設 A における結果を例に挙げる。分散が高い期間として、2010 年 7 月 15 日～2010 年 7 月 21 日、2011 年 12 月 29 日～2012 年 1 月 4 日に注目して述べる。

5.2.2 トピックの特徴量

トピックの特徴量について述べる。2010 年 7 月 15 日～2010 年 7 月 21 日、2011 年 12 月 29 日～2012 年 1 月 4 日における特徴量上位 25 件のトピックを表 5 に例として示す。

表を見ると、順位が大きく変動していないトピックが存在する。例えばトピック「ディズニー」が常に上位 5 位以内にある。これは施設 A が東京ディズニーリゾート（以下、TDL）のオフィシャルホテルであり、TDL に向かうユーザが利用しているためだと考えられる。また、飲食に関連するトピックが 10～25 位の一つ以上含まれている。これは施設 A が楽天トラベル飲食部門ランキングに入賞しているためだと考えられる。順位の変動が小さいトピックと特徴量が上位のトピックについて調べた結果、宿泊施設の特徴に関するトピックであった。各トピックと特徴量は施設について適切に捉えていると考えられる。

5.2.3 レビューの検出結果

レビューの検出精度と、検出されたレビューの例について説明する。

まず検出精度の比較結果について述べる。データセットの中

表 4 本実験で用いるトピック

キーワード	生成確率の高い上位 5 件の単語
カーテン	部屋 カーテン 残念 ビル 感じ
建物	建物 設備 感じ 自体 清掃
タオル	タオル スリッパ 浴衣 用意 バス
ビジネス 全国	ホテル ビジネス 宿泊 利用 サービス
きれい 快適	部屋 利用 きれい 快適 満足
マッサージ	マッサージ チェア 浴場 部屋 利用
男性	フロント 女性 男性 接客 従業
博多	便利 博多 立地 天神 利用
バス ユニット	バス ユニット ルーム タブ シャワー
トイレ	トイレ 風呂 部屋 洗面 便座
メニュー	朝食 メニュー バイキング 満足 種類
プラン 限定	プラン 宿泊 利用 限定 今回
きれい 綺麗	部屋 利用 きれい 快適 綺麗
出発	時間 チェックイン 朝食 出発 到着
確認 電話 フロント	フロント 対応 ホテル 電話 チェックイン
疲れ	浴場 温泉 疲れ 風呂 利用
露天風呂	風呂 食事 露天風呂 温泉 満足
インターネット	インターネット ネット 接続 無線 パソコン
料理	料理 食事 食材 温泉 風呂
ラウンジ	ラウンジ ホテル サービス レストラン バー
友人	宿泊 利用 予約 満足 今回
インター	便利 高速 近く インター 利用
体調	体調 病院 お世話 宿泊 感謝
月 次回	宿泊 月 利用 満足 大変
予約	予約 当日 宿泊 部屋 変更
お正月	お正月 大晦日 宿泊 元旦 正月
条件	立地 条件 最高 利用 駅前
徒歩	徒歩 便利 コンビニ 近く 立地
ランドリー	ランドリー コイン 洗濯 無料 乾燥
接客	部屋 満足 利用 大変 食事
是非	利用 部屋 満足 朝食 機会
ビジネス 普通	ホテル ビジネス 普通 部屋 感じ
廊下	部屋 廊下 夜中 ドア 防音
景色	部屋 最高 景色 露天風呂 食事
東京	利用 東京 ホテル 出張 立地
シャンプー	シャンプー アメニティ タオル 歯ブラシ ソープ
期待 正直	期待 部屋 ホテル 普通 宿泊
必要	部屋 必要 ホテル アメニティ 設備
温度	温度 エアコン 部屋 空調 調整
階段	エレベーター 階段 部屋 荷物 エレベータ
浴場	浴場 風呂 時間 入浴 残念
便利 近く	便利 近く 立地 利用 コンビニ
レンジ	レンジ 電子 マンション キッチン 洗濯
改善 検討	改善 お願い 今回 今後 検討
シャワー	シャワー 風呂 お湯 排水 トイレ
団体	団体 廊下 マナー ホテル 外国
プラン 今回	宿泊 満足 部屋 利用 今回
フロント 対応 笑顔	フロント 対応 スタッフ 女性 笑顔
対応 電話	対応 フロント チェックイン 電話 感謝
全体	部屋 満足 利用 きれい 快適

表 5 宿泊施設 A におけるトピックの特徴量上位 25 件の例

順位	2010 年 7 月 15 日～ 2010 年 7 月 21 日	2011 年 12 月 29 日～ 2012 年 1 月 4 日
1	ディズニー	ディズニー
2	対応 電話	笑顔 スタッフ
3	子供	子供
4	ツイン	混雑
5	シャワー	出発
6	誕生	チェック
7	レストラン	販売
8	駐車 無料	確認 電話 フロント
9	対応 フロント スタッフ	期待 正直
10	以前	レストラン
11	ベッド	前回
12	コンビニ 近く	アメニティ
13	プール	コンビニ スーパー
14	コーヒー	男性
15	前回	感謝
16	禁煙	種類
17	階段	雰囲気
18	月 次回	充実
19	楽天	電話 確認 事前
20	タオル	パン
21	電話 確認 事前	東京
22	バイキング	誕生
23	バス ユニット	感じ
24	リーズナブル	駐車 立体
25	料金	プラン 限定

で、レビュー件数が多い上位 10 施設を対象にレビューの検出を行った。実験環境は以下のように設定した。

- 対象とする施設：5.1 節で得られたデータセットにおいて、投稿件数が多い上位 10 施設
 - 正解データ：イベントが発生した期間の中で評価値が 2 以下のレビューまたは苦情や不満を述べているレビュー
 - 比較する指標
 - 順位相関
 - 全トピックにおける Jensen-Shannon 情報量
 - スコアの高い上位 30 件における Jensen-Shannon 情報量
- 評価期間によって総レビュー数、正解データ数は変化する。本研究では各評価期間における最低レビュー件数を 5 件としたため、検索レビュー数を 3 件とした。5.1 節の結果より、評価値が変動した要因は低い評価値を与えるレビューである可能性が高い。低い評価値を与えるレビューには苦情や不満を示す内容が記述される事を考慮し、正解データを設定した。

各宿泊施設においてイベントが発生した期間毎に適合率、平均適合率を求めてそれぞれを平均した。結果を表 6 に示す。

表 6 適合率、平均適合率による比較

	順位相関	Jensen-Shannon 情報量 (全トピック)	Jensen-Shannon 情報量 (上位 30 件)
適合率	0.487	0.34	0.593
平均適合率	0.692	0.533	0.81

適合率、平均適合率共にスコアの高い上位 30 件のトピックにおける Jensen-Shannon 情報量による検出精度が最も高い値を示した。適合率に比べて平均適合率が高いことから、検出されたレビューは比較的上位にあると考えられる。

次に、Jensen-Shannon 情報量が最も低いレビューと該当期間におけるスコアの分析結果について述べる。宿泊施設 A において分散が高い上位 5 つの期間は以下の通りとなった。

- 2010 年 7 月 15 日～2010 年 7 月 21 日
- 2010 年 7 月 18 日～2010 年 7 月 22 日
- 2010 年 7 月 20 日～2010 年 7 月 25 日
- 2011 年 3 月 29 日～2011 年 4 月 29 日
- 2011 年 12 月 29 日～2012 年 1 月 4 日

Jensen-Shannon 情報量が最も低いレビューが正解レビューではなかった期間は 2010 年 7 月 15 日～2010 年 7 月 21 日のみだった。2011 年 12 月 29 日～2012 年 1 月 4 日を正解レビューを示した期間の例、2010 年 7 月 15 日～2010 年 7 月 21 日不正解レビューを示した期間の例とする。該当期間におけるスコア上位 30 件のトピックを表 7 に示す。

表 7 宿泊施設 A におけるスコア上位 30 件の例

順位	2010 年 7 月 15 日～ 2010 年 7 月 21 日	2011 年 12 月 29 日～ 2012 年 1 月 4 日
1	ツイン	イン
2	コーヒー	東京
3	エコ	笑顔 スタッフ
4	誕生	男性
5	キー	エコ
6	禁煙	確認 電話 フロント
7	夕食	駐車 立体
8	ダブル	種類
9	駐車 無料	子供
10	バイキング	感謝
11	対応 電話	アメニティ
12	プール	期待 正直
13	レストラン	ビジネス 全国
14	アメニティ	コンビニ スーパー
15	子供	禁煙
16	マッサージ	ご飯
17	対応 丁寧 親切	パン
18	影響	お正月
19	和食	混雑
20	タオル	販売
21	駐車 立体	レストラン
22	料金	浴場
23	便利 近く	バイキング
24	プラン 限定	シャンプー
25	条件	新聞
26	テレビ	電話 確認 事前
27	是非	コーヒー
28	トイレ	アップ
29	友人	充実
30	メニュー	カーテン

2011 年 12 月 29 日～2012 年 1 月 4 日において検出されたレ

ビューを以下に示す。

- 投稿日時：2012 年 1 月 3 日 10 時 50 分
- 評価値：3
- 本文：「年末という時期を考えると、お安くおトクなプランだったと思います。（以前に泊まったときは、いわゆる閑散期だったのかもっと満足した記憶があるので）残念だった点は以下の 4 点です。・2 名で泊まったのにアメニティのヘアブラシ、ヘアバンドなどは 1 つしかなかったこと・シャンプー類はもう少し良いものを使ってほしい（髪がゴワついて仕方がなかった）・各階に製氷がなかったこと・食事（朝食）の際、会場に入るまでにかなり待ち、さらにバイキング形式で 1 列にしか並べず、かなり時間がかかったこと、またお皿を載せるトレイがあればよかった（お皿だと 2 枚しか持てない）お値段からすれば贅沢なわがままなのかもしれませんが、一流ホテルとして安価な部屋に泊まった客にも満足できるようにしていただければ嬉しいです。」

検出されたレビューには“アメニティ”や“混雑”に関連する内容が書かれており、スコアが高いトピックに関連するレビューが検出されたといえる。本論文では評価値が 2 以下のレビューまたは苦情や不満を示すレビューを正解レビューとしている。検出されたレビューは評価値が 3 だが、苦情や不満を示す正解レビューを検出できた。

2010 年 7 月 15 日～2010 年 7 月 21 日において検出されたレビューを以下に示す。

- 投稿日時：2010 年 7 月 18 日 15 時 20 分
- 評価値：5
- 本文：「ツインの予約でしたが、急遽ダブルに変更をお願いしましたら快くキングサイズのダブルのお部屋に交換してもらえました。朝食バイキングを昼食バイキングに変更できるプランもととても良かったです。昼間で寝て、ゆっくり朝昼食をランチバイキングで楽しみました。」

検出されたレビューには“ツイン”や“バイキング”に関連する内容が書かれており、スコアに関連するレビューを検出できている。レビューの評価値は 5 であり、内容も苦情や不満を示してはいない。“ツイン”のトピック寄与率が高いため、誤検出されたと考えられる。

6. ま と め

本研究では、レビューデータにおける評価値の変化に着目したイベント検出と要因分析の手法を提案した。具体的には、レビュー対称毎に評価期間を設定してイベントが発生した時期を検出し、その要因を分析する手法を提案した。実験では順位相関による検出結果と Jensen-Shannon 情報量による検出結果を比較する事で検出精度の評価を行った。また検出されたレビューの内容を分析する事で、手法の有効性を示した。

評価値が変動した要因を分析した結果、苦情や不満を示すレビューによって変動するケースが多い事が分かった。苦情や不満を示すレビューを検出する手法を用いる事で、要因を示すレ

ビューの検出精度が向上すると考えられる。

謝 辞

本研究で使用したレビューデータは、楽天データ公開によるものである。ここに記して謝意を表す。本研究の一部は科学研究費補助金若手 B(23700102) による。

文 献

- [1] 楽天データ公開. <http://rit.rakuten.co.jp/rdr/index.html/> .
- [2] 乾孝司, 奥村学. テキストを対象とした評価情報の分析に関する研究動向. 自然言語処理, Vol.13, No.3 2006.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.2 2005.
- [4] 辻井康一, 津田和彦. テキストマイニングを用いた宿泊レビューからの注目情報抽出手法. 情報処理学会デジタルプラクティス, Vol.3, No.4, 2012.
- [5] 藤井絵美子, 上野剛, 中本政一, 東高宏, 加藤直樹, 羽室行信. ホテル業界における口コミ情報に基づいた顧客満足度予測モデルの構築とポジショニング分析. WebDB Forum 2011.
- [6] 北島理沙, 小林一郎. 文書内の事象を対象にした潜在的ディリクレ配分法による要約. DEIM Forum 2011.
- [7] L.Henning. Topic-based Multi-Document Summarization with Probabilistic Latent Semantic Analysis. RANLP 2009.
- [8] Ivan Titov, Ryan McDonald. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. ACL HLT 2008.
- [9] 山岸祐己, 斎藤和巳, 大久保誠也. レビュー時系列データからの分割統治による変化点検出法. JSAI 2012.
- [10] Y.W.Teh, M.I.Jordan, M.J.Beal, D.M.Blei. Hierarchical Dirichlet Processes. JASA 2006.
- [11] 嘉村亘太, 黄宏軒, 川越恭二. トピックモデルにおけるトピック表現語導出. DEIM 2011.
- [12] MeCab: Yet Another Part-of-Speech and Morphological Analyzer.
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>.
- [13] MALLET: A Machine Learning for Language Toolkit. Andrew Kachites McCallum. <http://mallet.cs.umass.edu> 2002.