

店舗レビューには何が書かれているか？—調査及びその自動分類—

新里 圭司

楽天技術研究所

keiji.shinzato@rakuten.com

1 はじめに

オンラインショッピングでは出店者（以下、店舗と呼ぶ）と顔を合わせずに商品を購入することになるため、店舗とのやりとりは顧客満足度を左右する重要な要因となる。商品の購入を検討しているユーザに対して、商品を扱っている店舗が「どのような店舗か」という情報を提供するため、楽天市場では店舗に対するレビューの投稿・閲覧ができるようになっている。店舗レビューの例を図1に示す。自由記述以外に6つの観点（品揃え、情報量、決済方法、スタッフの対応、梱包、配送）に対する5段階評価（5が最高、1が最低）がついているが、評価基準がユーザによって異なるのに加え、ここに挙がっている観点以外の情報も自由記述に含まれるため、店舗を詳細に調べようとする場合は目を通す必要がある。レビュー内の各文をその内容に応じて分類することができれば、店舗の良い点、悪い点などユーザが知りたい情報へ効率良くアクセスできるようになり、今までであったユーザの負担を軽減することが期待できる。オンラインショッピングサイト運営側の点からも、顧客満足度を下げている店舗をコメントに記載されたクレームとともに発見できるようになるためメリットは大きい。

本稿では、このような情報アクセス技術の開発を目的に、店舗レビュー中に記述されているトピックを調査したのでその結果について報告する。さらにレビュー内の各文をトピックおよびその評価極性（肯定、否定）に分類する方法についても述べる。

2 店舗レビューには何が書かれているか？

調査にあたり、まず楽天市場の店舗レビューを評価が高いものと低いものに分類した。具体的には、各店舗レビューに含まれる6つの評価観点に対する評価値の平均を求め、値が4.5以上のものを高評価、2.0以下のものを低評価レビューとした。続いてレビューを「\n」「。」「!」「?」「」「...」「..」「...」を手がかりに文



図1: 楽天市場における店舗レビューの例。

に区切った¹。そして、高評価、低評価に分類されたレビューのうち、3文以上含むものを無作為に50件ずつ選びだし、それらに記述されている内容（トピック）を調査した。対象となったレビュー100件には487文含まれていた。

調査の結果得られたトピックを表1に示す。実際の調査ではPaymentに関する記述は見当たらなかったが、決済方法に関する観点が店舗レビューに含まれるため追加した。表よりメール・電話などの店舗とのやりとり、店舗に対する評価、再度購入したいと思ったかどうか、キャンセル・返品時の対応など、先述した6つの評価観点では明確に捉えられていないトピックについてレビュー中で言及されていることがわかる。また店舗レビューではあるが、商品やその価格に関する言及もあることがわかる。

3 トピックおよび極性の自動分類

表1のトピックをもとに、店舗レビュー1,510件（5,266文）に対して人手でアノテーションを行い、このデータに基づいてトピックおよびその評価極性を分類するモデルを構築した。本節ではこれらについて述べる。

3.1 学習データの作成

2節同様、店舗レビューを高評価および低評価に分類し、高評価レビューから780件、低評価レビューから730件無作為に選びだしアノテーションの対象とした。

¹ただし「」内に出現する場合は区切らない。また(1)直前の文が「:」もしくは「、」で終了している場合、または(2)直線の文が「?」又は「!」で終了し、当該文が「が」「とても」「とりあ」「ところ」で始まっている場合は当該文を直前の文に連結した。

表 1: 店舗レビューで言及されるトピックおよび各トピックの評価極性の分布 .

トピック	説明 () 内はトピックに該当する文の例 .	事例数 (レビュー 1,510 件 , 5,266 文)			
		肯定	中立	否定	合計
Shipping	発送 / 送料 / 配送業者 (商品が、あっという間に手元にきました。)	380	171	510	1,061
Service	ショップ、スタッフの対応・態度 (要望に応えて頂いたうえに迅速な発送で感激でした。)	431	39	562	1,032
Cumunication	メール、電話、FAX などショップとのやり取り (連絡、配送ともに迅速に対応して頂きました。)	102	188	644	934
Merchant	ショップに対する評価・感謝、苦情など (お電話での対応もよく、迅速に進めて頂き、信頼できました。)	298	30	567	895
Item	商品、ブランド品の場合、その保存箱等も商品とみなす (境目がチープさを強調しています。)	235	141	279	655
Repeat	リピートしたいかどうか (今後もお世話になります。)	391	6	167	564
Package	梱包 (箱が、めちゃくちゃ変形してました。)	208	38	98	344
Stock/variety	商品の在庫 / パラエティ (外国のビールを多く取り扱っていて、値段も良心的です。)	63	26	196	285
Information	商品販売ページに掲載されている情報 (注文した商品は写真と全然違います)	9	25	192	226
Cancel/return	キャンセル / 返品 (サイズが合わないのに交換してくれません。)	6	68	134	208
Price	商品の価格 (商品はもっと安くなったらありがたい…)	117	28	28	173
Rakuten	楽天自体のサービス、システム、対応など (楽天にも責任があると思います。)	3	6	21	30
Payment	支払方法 (クレカが使えるので、決済 5)	3	1	5	9
Other	上記以外、購入の背景や感想など (謝恩会の委員の方たちへのお礼として購入しました。)	121	356	220	697
合計		2,367	1,123	3,623	7,113

調査に用いたレビュー 100 件を使って訓練した後、1 名のアノテータに 5,266 文に対して該当する全てのトピックおよびその評価極性 (肯定, 中立, 否定) を付与するよう依頼した。Other があるため各文には必ず 1 つ以上のトピックおよびその評価極性が付与されることに注意されたい。また「驚きました!」のように文単体ではトピック・評価極性の同定が難しい場合は、前後の文脈を利用するように指示した。そのため、同じ文であっても前後の文脈により付与されるトピック・評価極性が異なる場合がある。

5,266 文における各トピックおよびそれらの評価極性の分布を表 1 に示す。1 文に対して複数のトピック、評価極性が付与されることがあるため実際の文数よりもアノテーションされた文の総数は多くなっている。Communication, Merchant, Repeat, Information, Price などのトピックは評価極性の分布に偏りが見られる。

3.2 分類モデルの構築

3.1 節で構築したトピックおよび極性タグ付きデータを使いトピック分類モデルおよび評価極性分類モデルを構築する。表 1 からわかるように、評価極性分類モデルの構築に十分な量の事例数が確保できないトピックがあるため、トピック分類および評価極性分類のモ

デル構築は独立に行った。

分類にあたり、文ごとに分類を行うモデル (文単体モデル)、レビューに含まれる各文の分類を系列ラベリング問題とみなし、レビューごとに分類を行うモデル (文章モデル) の 2 種類を考えた。文章モデルを設けた理由は、前後の文脈を利用しないとトピック・評価極性の同定が難しい曖昧な文があるためである。

3.2.1 トピックの分類

より高い精度 (Precision) を得るため、文単体モデルおよび文章モデルの両方を構築し、両モデルにより推定されたトピックが同じ場合のみトピックを出力した。そのため、現状では 1 文に対して 1 つのトピックしか推定できない。複数のトピックを推定する方法については今後検討する予定である。

文単体モデルおよび文章モデルでは以下に示す共通の素性を用いた。

単語素性: 品詞の大分類が名詞 (細分類が非自立のものは除く)、形容詞、動詞、感動詞、副詞となる単語の原型。単語の直前、もしくは後方に否定を表す語 (「不～」や「～ない」等) が続く場合は、否定を表すフラグを付与し、肯定の場合と区別した。

辞書素性: 人手で構築した辞書 (表 2) の要素と一致

表 2: トピック分類用辞書の例 (全 171 件) .

トピックラベル	要素の例
Shipping	発送日, うけとる, 冷凍便, 納品, ゆうパック, クロネコ, 送料
Service	対応, 気遣い, 態度, 心遣い, 対応, スタッフ
Communication	返信, 連絡, 返答, FAX, TEL, メール
Merchant	業者, ショップ, 店, おまけ, 店舗
Item	カラー, 傷, 写真, 納品書, 品物, 色, キズ
Repeat	リピート, お世話, 利用, 機会, リピ
Package	梱包, エアクッション, ブチブチ, 包装紙
Stock/variety	売りきれ, 品薄, 取り扱う, 品ウス, 売切れ
Information	嘘, サイト, ウソ, ページ, 情報, 記載
Cancel/return	返品, 送り返す, CANCEL, 注文ミス, 交換
Price	円, 値段, 金額, 価格, 万円, 0円, 千円

する形態素 (列) が文中に含まれていれば, 要素に対応するラベルを素性とする .

3.2.2 評価極性の分類

トピックが推定された文に対してのみ分類を行う . 同じ評価極性の文が続けて現れやすい文脈一貫性 [1] を利用するため, 評価極性の判定には文章モデルを採用した . 文単体モデルと併用しなかった理由は, 4.1 節で示すように文章モデルだけである程度の精度が達成できたこと, および全体としてのカバレッジがさらに下がることを避けるためである . 評価極性の分類には以下の素性を用いた .

単語素性: トピック分類と同じ .

接続詞素性: 「でも」「ただ」「しかし」「だけど」で文が始まっているかどうか .

評価極性素性: 文内に含まれるポジティブな語, ネガティブな語の数 .

評価極性素性を抽出するにあたり, 次の方法で単語に対してポジティブ, もしくはネガティブの度合いを計算した . まず 2016 年に楽天に投稿された店舗レビューの中から, 5 文以上含む高評価, 低評価レビューを無作為に 5 万件ずつ集めた . 文分割方法および高評価, 低評価の判定方法は 2 節と同じである . 次に集められたレビューに含まれる単語を対象に, 高評価レビュー中での文書頻度 df_{pos} , 低評価レビュー中での文書頻度 df_{neg} を算出し, $df_{pos} > df_{neg}$ であればポジティブ, それ以外はネガティブとして評価極性を決定した .

4 実験

4.1 性能評価

3.1 節で作成したデータをもとに 10 分割交差検定を行い手法の性能を評価した . 評価の際, Rakuten および

表 3: 実験結果 . P は精度, R は再現率を表す .

トピック	トピック分類			評価極性分類	
	P	R	F1	P_{pos}	P_{neg}
Shipping	0.846	0.756	0.798	0.853	0.860
Service	0.827	0.552	0.662	0.943	0.900
Communication	0.915	0.742	0.819	0.833	0.910
Merchant	0.735	0.674	0.703	0.941	0.927
Item	0.697	0.388	0.498	0.762	0.696
Repeat	0.898	0.864	0.880	0.939	0.943
Package	0.926	0.714	0.806	0.824	0.778
Stock/variety	0.702	0.446	0.545	0.750	0.903
Information	0.762	0.372	0.500	0.667	0.970
Cancel/return	0.778	0.625	0.693	None	0.854
Price	0.926	0.568	0.704	0.944	0.444
Other	0.549	0.344	0.423	0.667	0.731
マクロ平均	0.797	0.669	0.728	0.829	0.826
マイクロ平均	0.797	0.619	0.697	0.883	0.873



Training data₁: Communication → Shipping → Repeat
 Training data₂: Communication → Package → Repeat

図 2: 複数のトピックが付与されたデータの展開 .

Payment は事例数が少ないため学習が困難と判断し, これらは Other として扱った . 文単体モデルの学習には opal² を, 文章モデルの学習には CRFSuite³ を用いた . opal で学習する際は 2 次の多項式カーネルを用い, 他のパラメータはデフォルトを利用し, CRFSuite での学習は全てデフォルトのパラメータを利用した . 素性を抽出する際は前処理として MeCab⁴ および NAIST Japanese Dictionary⁵ を使って文を形態素解析した .

モデルを学習する際, 複数のトピック・評価極性を持つ文は, その全てのトピック・評価極性の訓練事例とした . 例えば, 図 2 は複数のトピックを持つ文を含むレビューの例である . 文単体モデルでは s_2 を Package, Shipping の事例として扱った . 一方で文章モデルの場合は, ラティス構造中の全てのパスを学習データとして利用した⁶ .

3.2.1 節で述べたように提案手法は 1 つの文に対して複数のトピックおよび極性を付与できない . そこで今回は, トピックが 1 つの文だけを評価対象とした . 評価結果を表 3 に示す . トピック分類は F1 値で 73 ポイント, 評価極性の分類は肯定, 否定共に 83% 程度の精度で行えていることがわかる (マクロ平均) .

²<http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/opal/>

³<http://www.chokkan.org/software/crfsuite/>

⁴<http://taku910.github.io/mecab/>

⁵<https://ja.osdn.net/projects/naist-jdic/>

⁶評価極性分類の際も同様に全てパスを利用した .

表 4: エラーのタイプおよびその事例数.

タイプ	事例数
分類に必要な素性が得られていない	30
文単体では分類が難しい	21
文の記述内容の把握が必要	12
アノテーションの誤り	4
その他	2

4.2 トピック分類の誤り分析

無作為に選んだ 1,358 件のレビュー (4,740 文) で学習を行い, 残り 152 件 (537 文) をトピック分類した際の false-positive の分析を行った. 分析の対象となった事例は 63 件であり, 各事例には 1 つ以上のエラータイプを付与した.

分析の結果を表 4 に示す. 最も数が多かったタイプは「分類に必要な素性が得られていない」で 30 件あった. 例えば「欲しい色とサイズもあって、良かったです。」は Stock/variety に分類されるべきであるが, 提案手法は Item に分類していた. 抽出された素性と各トピックの関連を調べたところ, Item と関連の強い素性が多く, Stock/variety と関連のある素性はなかった. また商品に関する記述を含む文は Item と関連の強い素性が抽出されていない傾向にあった. これは, 楽天で多様な商品を扱っていること, 店舗レビューの性質上, 商品に関する記述はメインでないことから十分な量の訓練事例が得られていないためと考えられる.

次に多かったタイプは「文単体では分類が難しい」であり, 21 件が該当した. 例えば文「あきれます。」は商品にも店舗にも当てはまる表現であり, この文を正しいトピックに分類するためには, この文の前後にどのような記述があるのかを捕らえる必要がある. 現状では文単体モデル, 文章モデルともに明確な素性として前後の文脈を用いていない. このタイプのエラーを減らすには, 何らかの方法で前後の文脈の内容を取り込む素性が必要がある.

3 番目に多かったタイプは「文の記述内容の把握が必要」であり, 12 件が該当する. 例えば「親戚宅への訪問土産に購入しました。」のような購入背景, 「重さのあるものも家の玄関に届くのでとっても助かっています。」のような感想がこのタイプに当てはまる. これらを正しく分類するためには, 学習データを増やす, もしくは現在 Other として分類しているこれらの文をより細かく分けて学習するなどが挙げられる.

また「アノテーションの誤り」も確認された. 例えば文「別に目立つほど悪い点は無かったです。」は Other が付与されていたが, 提案手法は Item に分類していた. レビューを確認するとこれは商品についての記述

であり Item が付与されることが妥当である. このようなエラーを減らすには複数人のアノテータにアノテーションを依頼し多数決をとる, または分類基準を精緻にするなどが考えられる.

5 関連研究

乾ら [3], 浅野ら [5] は宿泊施設レビューを対象に文レベルでのトピック分類および評価極性分類を行っている. 安藤ら [2] は商品レビューを対象に, そこに記述されている内容を調査し, 23 種類のカテゴリを提案している. いずれの研究もレビューに記載されている内容の調査・(自動)分類という点で本研究と類似しているが, 対象となるレビューが宿泊施設, 商品であり本研究で対象とした店舗と異なる.

山下ら [4] は商品レビューに含まれる店舗に関する記述を検出する手法を提案している. しかしながら, 店舗に関する記述を特定するだけで, それらのトピックや評価極性の分類は行っていない.

6 おわりに

本稿では, オンラインショッピングサイト出店者 (店舗) に対するレビューを対象に, どのようなトピックについて言及されているのかを明らかにした. また明らかとなったトピックをもとに, レビュー中の各文に対して人手でアノテーションを行い, 機械学習を使ってレビュー中の各文をトピックおよびその評価極性に分類することを行った. その結果トピック分類を 73 ポイント (F1 値) で, 評価極性分類を 83% (精度) で行えることを示した. しかしながら, 提案手法では 1 文に対して 1 つのトピック・評価極性しか推定できず, 複数への対応は今後の課題である.

参考文献

- [1] Hiroshi Kanayama and Tetsuya Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of EMNLP 2006*, pages 355–363, 2006.
- [2] 安藤 まや, 関根 聡. レビューには何が書かれているのか? In *ALAGIN & NLP 若手の会合同シンポジウム*, 2013.
- [3] 乾 孝司, 梅澤 佑介, 山本 幹雄. 評価表現と文脈一貫性を利用した教師データ自動生成によるクレーム検出. *自然言語処理*, 20(5):683–706, 2013.
- [4] 山下達雄, 東野進一. 商品レビューに含まれるストア言及の抽出. In *情報処理学会 第 78 回全国大会講演論文集*, 2016.
- [5] 山本幹雄 浅野翔太, 乾孝司. 談話役割に基づくクラス制約規則を利用したレビュー文の意見分類. In *言語処理学会第 20 回年次大会発表論文集*, pages 880–883, 2014.