

商品紹介文改善のための 紹介文およびレビュー集合における評価視点の整理

角田 孝昭*

筑波大学大学院 システム情報工学研究科

tsunoda@mibel.cs.tsukuba.ac.jp

Zofia Stankiewicz

関根 聡

楽天株式会社 楽天技術研究所

{zofia.stankiewicz,satoshi.b.sekine}@mail.rakuten.com

1 はじめに

大量の商品やサービス（以下、まとめて商品と呼ぶ）を販売しているオンラインショッピングサイトでは、商品検索ページなどで一度に多くの商品をユーザに提示できるようにするため、各商品を紹介できるスペースを限っている場合が多い。そのため、商品販売者は各商品の紹介文を限られた文字数の下で可能な限り魅力的に記述することが求められる。しかし、実際には販売者が紹介文で訴求する特長とユーザが真に求める特長は必ずしも一致せず、紹介文によっては本来備えている魅力がユーザに伝わらない場合がある。

このような販売者とユーザ間のギャップを埋めるため、本研究では商品紹介文では言及されていないが特徴的な商品特長を自動で抽出するタスクを提案する。この情報は、販売者にとって商品紹介文の改善に有用であるほか、検索エンジン最適化 (SEO) などの広告戦略の見直しにも活用可能である。更に、ユーザにとっても購買判断にあたって有用な情報である。

タスクの概要について、具体的な例（宿泊サービス）を用いて説明する（図1）。例えば、紹介文では宿泊施設からの「景観」についてのみ触れているが、分析の結果「景観」「風呂」「食事」「外装・内装」の順に特徴的と言う情報があれば、紹介文になかった「風呂」「食事」について、更に紹介文のスペース次第で「外装・内装」についても紹介文に加える、と言った戦略を立てることができる。

実際に特徴的な商品特長を判断するための材料として、本研究では商品のレビュー集合に注目する。レビューにはユーザが特徴的と感じた特長について記述されており、中には商品紹介文にはない特長についての言及も数多く存在する。我々はこのような商品レビュー集合を活用することで、商品紹介文で触れられていなくてもユーザにとっては特徴的である商品特長を抽出する。

商品紹介文

全ての部屋から見えるオホーツク海に沈む夕陽は絶景



特長	対応するレビュー文の例
景観	早めに到着ということもあり、オホーツク海に沈む夕日をバッチリ見れました。
風呂	温泉は掛け流しで気持ちがよく、2ヶ所ある時間制貸切露天風呂は最高でした。
食事	お料理はオイルフォンデュ（いわゆる自分で調理する天ぷら）やら ...
外装・内装	建物も、木のぬくもりを感じられる雰囲気の良いもので、...

図1: 本研究で提案するタスクの概要（各文は楽天トラベルより引用）。風呂, 食事, 外装・内装を提示する

本研究ではこのような商品特長を抽出するために、従来の商品特長の構造化手法・商品特長のランキング手法を組み合わせた手法を提案する。加えて、タスクの特性に基づいて商品特長の特徴度計算及び商品特長の選択に工夫を加える。

実際に楽天トラベル¹（この場合「宿泊サービス」が商品となる）のデータを用いた実験の結果、提案手法により特徴的かつ紹介文にない商品特長の抽出が実現したことを示す。

2 関連研究

本章では本研究と関連の深い研究、特に評価視点 (opinion features) の抽出や整理を対象とした研究について述べる。

まず、評価視点の抽出を行う研究に注目する。具体的な評価の対象を明らかにすることは評判分析において最も重要なタスクの一つであるため、現在までに様々な研究が行われている [Hu and Liu, 2004] [小林他, 2005] [Xu et al., 2013]。近年ではこれらの研究を更に発展させ、列挙した評価視点を何らかの観点で整理す

*本研究は、著者が楽天技術研究所ニューヨークのインターンシップの一環で行ったものである。

¹楽天トラベル: <http://travel.rakuten.co.jp/>

る研究も進められている。本研究も、特徴的かつ商品紹介文で言及されていない評価視点のみを提示する点で、評価視点の整理を目的とした研究と言える。

本研究では、従来研究の評価視点の構造化手法・評価視点のランキング手法を組み合わせることでタスクを実現する。それぞれの従来研究について以下で説明する。

評価視点を構造化する研究として、Zhai らによる各評価視点を予め人手で定義した評価視点グループに対応付ける研究がある [Zhai et al., 2010]。Zhai らは、Nigam らによる単純ベイズアルゴリズムを拡張した手法 [Nigam et al., 2000] を応用することで、半教師あり学習の枠組みで評価視点の対応付けを実現している。

評価視点をランキングする研究として、乾らの研究や Yu らの研究が挙げられる [乾他, 2013] [Yu et al., 2011]。中でも、乾らは他商品レビュー集合と比較した際に、特に注目商品のレビュー集合で特有な評価視点を特徴的評価視点と位置付け、他商品との比較と言う観点に基づいた相対的特徴に基づいてランキングを行う。具体的なランキング手法は、注目する商品と評価視点の出現が依存していると仮定した場合にデータ全体が観測される確率と、独立していると仮定した場合の確率の比 (対数尤度比) に基づいている。

3 提案手法

本研究では従来研究に倣い、商品特長を評価視点と言う概念で扱う。ここで、評価視点とは商品の一部分や属性を指す²。

本研究で抽出したい評価視点の条件の一つは、評価視点の商品紹介文で言及されていないことである。本研究ではそのような評価視点の抽出を、評価視点を「評価視点グループ」と言うより粗なレベルで扱うことで実現する (3.1 節)。

併せて、評価視点グループをどのようにスコア付けするか (3.2 節)、スコア付けした評価視点グループをどのように選択するか (3.3 節)、と言う点に工夫を加える事で性能の向上を図る。

3.1 評価視点グループへの対応付け

商品紹介文に述べられていない評価視点を抽出する最も単純な方法に、商品紹介文及びレビュー集合から言及されている評価視点表現を抽出して、それらの差分を提示する手法がある。しかし、この手法によると例えば紹介文で「アクセス抜群のホテル」、レビュー文書で「交通の便が良い」「駅から近い」と述べられていた場合、「アクセス」と「交通の便」「駅」がマッチせず、紹介文にない評価視点の候補として「交通の

²例えば「宿泊サービス」を例に説明すると、評価視点には「朝食」「露天風呂」「交通の便」「フロントの対応」などが対応する。

便」「駅」などが提示される可能性がある³。いずれの文も宿泊施設の立地について述べている点では同一であるため、「交通の便」「駅」を提示してもあまり有用ではない。

このような問題に対処するため、本研究では評価視点を評価視点グループと言う粗なレベルに対応付けた上で、商品紹介文及びレビュー集合のマッチを行う手法を提案する。例えば「宿泊サービス」の場合、予め「立地」「食事」「風呂」と言った評価視点グループを用意し、各評価視点を評価視点グループへ対応付けた上でマッチを行う。これにより、評価視点「アクセス」「交通の便」「駅」は「立地」に対応付けられるため、これらは新たな評価視点として提示されなくなる。

具体的な手順は次の通りである。まず最初に、全ての評価視点を評価視点グループに対応付けた辞書を作成する。このような辞書を人手で作成するには非常に高いコストを要するが、我々は Zhai らの手法を援用し、半教師あり学習により低コストで作成した [Zhai et al., 2010]。次に、商品紹介文及びレビュー集合から抽出した各評価視点に対し、辞書を参照して評価視点グループへの対応付けを行う。この対応付けにより、商品紹介文及びレビュー集合で言及されている評価視点を評価視点グループのレベルで比較できる。

3.2 評価視点グループの特徴度スコア付け

商品紹介文にない評価視点グループのみに限っても、レビュー文書の規模が大きい場合は多くの評価視点グループが抽出されてしまう。そのため、何らかの基準に基づいてスコア付けを行い、高いスコアを得た評価視点グループのみを提示することが望ましい。

そこで、我々は特徴的評価視点と言う基準に基づいて評価視点のスコア付けを行う乾らの先行研究 [乾他, 2013] を援用する。具体的には、乾らの手法における「評価視点の出現頻度」を「評価視点グループの出現頻度」に置き換えることで、評価視点グループのスコア付けを実現する。すなわち、各商品ごとに商品と評価視点グループが独立している度合いを対数尤度比により計算し、特徴的な評価視点グループを特定する。

加えて、我々は評価視点グループの出現回数を単語単位やレビュー単位で数えることにする。これは、単純に単語単位で数えた場合、すなわち「各評価視点グループに属する全評価視点表現の出現回数の総和」を評価視点グループの出現回数とする場合⁴、評価視点表現の出現に偏りのあるレビューに影響される問題が生じ

³この際、同義語をまとめて扱うことで「アクセス」「交通の便」などを同一に扱える可能性はあるが、「駅」も同一に扱うためには同義語レベルよりも粗なレベルで扱う必要がある。

⁴この場合は、乾らの研究 [乾他, 2013] において異表記をまとめて上げる際に、同じ評価視点グループにある全ての評価視点表現を異表記とみなす場合に相当する。

るためである。例えば、レビューが他のレビューと比較して特定の評価視点グループについて数多く言及している場合（具体的には「食事」について長文で苦情を述べる場合などが該当する）、出現頻度に大きく影響を与えてしまう。これに対し、同じ文または同じレビューにおいて、ある評価視点グループに属する評価視点表現が2回以上出現している場合でも1回とみなして出現頻度を数えることで、このような偏りのあるレビューの影響を抑えられると期待できる。

3.3 評価視点グループの選択

最終的に提示する評価視点グループの選択は、評価視点グループのスコア付けに基いて行う。この選択は、単純にはスコアが予め定めた閾値以上であれば選択すると言った手法でも実現可能である。

しかし、閾値による選択では評価視点グループが一つも選択されない商品や、過剰に選択されてしまう商品が出てきてしまう問題が起りうる。実際には、例えば宿泊施設の「立地」が悪くてもその分「価格」が安いなど評価視点グループ間がトレードオフの関係にある場合も多いため、特徴的な評価視点グループの数は限られている。

そこで、本研究では各商品に対して N-best により評価視点グループを選択する。これにより、スコアの絶対値によらずに一定数の評価視点グループが選択されるため、この問題を回避することができる。

4 実験

4.1 実験条件

実験に用いるデータには、楽天トラベルにおける宿泊施設の紹介文及びレビュー集合を用いた。紹介文のデータとして宿泊施設の詳細情報ページに掲載されている情報を、レビュー集合のデータとして「楽天データ公開⁵」に収録されているデータを用いた。このうち、レビュー件数が10件から1000件の施設を選択し、13,664施設、2,254,307レビューを得た。前処理に必要な単語分割は MeCab 0.996⁶ により行い、文分割は簡易的なヒューリスティックスにより行った。

評価視点表現-評価視点グループの対応辞書は、Zhaiらによる半教師あり学習の手法 [Zhai et al., 2010] を用いて作成した。この際、予め12種類の評価視点グループ⁷を定義し、出現頻度が高かった名詞281語を人手

で対応付けたデータをシードとした。対応付ける評価視点表現の候補としては、少なくとも全体の1%以上の施設のレビュー集合において出現した名詞及び名詞句⁸のうち、固有表現などを排除した9,844語を用いた。作成した辞書に対して文単位での評価⁹を行ったところ、適合率51.87%、再現率86.07%を得ている。

提案手法は、評価視点グループの頻度を文単位・レビュー単位で数えて対数尤度比によりスコア付けを行い、N-bestで選択した手法 (S-LLR, R-LLR) である。比較手法として、単純に評価視点グループの頻度をスコアとする手法 (R-TF)、各施設のレビュー集合における評価視点グループの相対頻度のTスコアをスコアとする手法 (R-Tscore) を用いる。また、スコア付け・選択の工夫を行った手法と比較するため、評価視点グループを単語単位で数えて対数尤度比を計算した手法 (W-LLR)、R-LLRの計算後に閾値に基いて選択した手法 (R-LLR/ABS) とも比較する。

評価は、模範的な施設紹介文で言及されている評価視点グループを正解とみなして行う。この理由は、各施設の特徴的な評価視点を総合的な側面から評価して順序を付けるのは、非常に高いコストを要するためである。これに対し、模範的な施設紹介文ではほぼ確実に特徴的な評価視点について言及している上、正解データも容易に作成できると言う利点がある。これら紹介文で言及されている評価視点とレビュー集合から抽出した特徴的な評価視点が一一致していれば、魅力的でない紹介文を持つ施設に対しても模範的な施設紹介文へと改善するための情報を提示すると期待できる。具体的なデータには、無作為に選んだ135件の施設のうち、人手で「紹介文に魅力を感じない」と判断した施設を除いた116件を模範的な施設紹介文とし、人手で評価視点グループを付与したデータを用いる。

評価指標には、適合率・再現率を用いる。

4.2 実験結果

実験結果及び考察を示す。なお、本タスクでは {2,3}-best 程度における性能が重要である。なぜなら、あまりに多数の評価視点を候補として提示した場合、タスクの目的である「限られた文量の紹介文に追加するための評価視点の提示」が達成できないためである。

まず、特徴的な評価視点グループのスコア付けにも対数尤度比が有用かを検証する。図2に、評価視点グループの出現頻度カウント方法をレビュー単位に固定して、スコア付け手法のみを変えた結果を示す。図を

⁵楽天データ公開: <http://rit.rakuten.co.jp/rdr/>。データは2012年当時の公開データを用いた。

⁶MeCab: Yet Another Japanese Dependency Structure Analyzer: <https://code.google.com/p/mecab/>

⁷サービス, 立地, 景観, 外装・内装, 部屋, 室内設備, 館内設備, アメニティ, 風呂, 食事, 価格, ウェブの12種類。

⁸名詞の連続 (展望+露天風呂, など)、AのB (スタッフの対応, など)、の二種類。

⁹辞書の評価用には320文を無作為抽出した。具体的な評価は、文中に出現した各評価視点を評価視点グループに対応付けた結果を自動付与された文のラベルとし、人手の結果と比較することで行う。なお、マルチラベルも許容する。

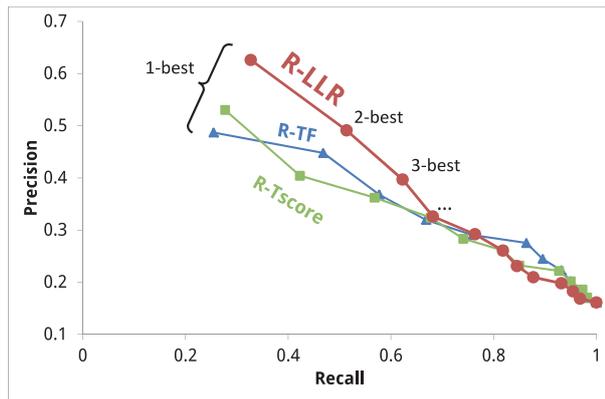


図 2: 比較手法と対数尤度比による手法の比較

見ると、概ね対数尤度比によるスコア付け (R-LLR) が高い性能を示しており、評価視点グループのスコア付けにも対数尤度比が有効であることが分かる。

次に、評価視点グループの出現頻度カウント方法を変えた場合の性能を比較する。図 3 に、スコア関数を対数尤度比に固定して、出現頻度のカウント方法と評価視点グループの選択方法を変えた結果を示す。図を見ると、文単位 (S-LLR) 又はレビュー単位 (R-LLR) で評価視点グループの出現頻度をカウントした場合の方が、単語単位で数えた場合 (W-LLR) よりも低～中再現率でやや高い性能を示している。しかし、性能差は大きくないため別の観点から改善を行う必要があると考えている。

最後に、評価視点グループの選択方法を変えた場合の性能を図 3 で比較する。図を見ると N-best で選択した場合 (R-LLR) の方が、閾値に絶対値を設定して選択した場合 (R-LLR/ABS) よりも低～中再現率で高い性能となっており、概ね N-best 選択の方が有効であることが分かる。

最も高い性能であった R-LLR の性能を見ると、3-best で再現率・適合率がそれぞれ 62.27%, 39.71% と、ある程度紹介文に相応しい評価視点グループを提示できている。より実用に近づけるためには、特に適合率の改善があると考えている。

5 おわりに

本研究では、商品紹介文の改善を主な目的とし、商品紹介文にない特徴的な評価視点を抽出する手法を提案した。

今後の課題として、以下の 2 点がある。

まず、どのような評価視点が商品紹介文に相応しいかの事前知識を利用できることが望ましい。例えば、紹介文全体の傾向として「外装・内装」よりも「食事」の方が書かれやすいと言う事前知識があれば、より「食事」を提示するように改良できる。更に、この事

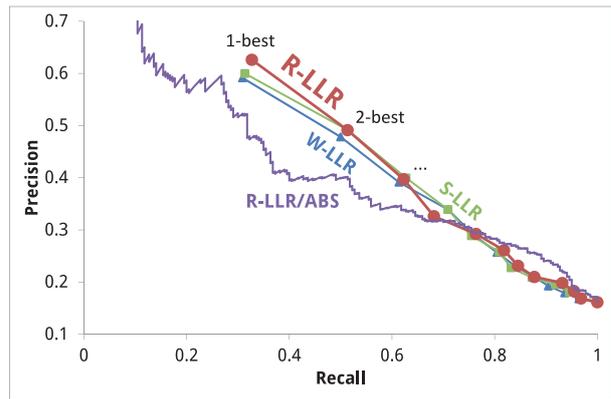


図 3: 頻度カウント方法及び候補選択方法の比較

前知識を逆に利用し、他の商品紹介文と差を付けられるような評価視点が必要な場合は「食事」を下位に順位付けると言った応用も可能になる。

また、場合によって細かいレベルでの情報も併用できることが望ましい。基本的には、我々の提案したタスクでは粗なレベルでのマッチがより適切であるが、より細かいレベルの情報も必要となるケースも存在する (例えば「カニ」「鍋」はいずれも評価視点グループ「食事」に対応するが、これらを陽に区別したい場合など)。このようなケースに備え、場合によって評価視点と評価視点グループ双方のレベルを考慮できればより有用になると考えている。

謝辞

本研究には楽天データ公開において公開されたデータを使用させて頂きました。貴重なデータを公開して頂きました楽天株式会社に感謝致します。

参考文献

- Minqing Hu and Bing Liu, 2004. "Mining opinion features in customer reviews," in *Proceedings of the 19th national conference on Artificial intelligence*, pp. 755–760.
- Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell, 2000. "Text Classification from Labeled and Unlabeled Documents using EM," *Machine Learning*, Vol. 39, No. 2-3, pp. 103–134.
- Liheng Xu, Kang Liu, Siwei Lai, Yubo Chen, and Jun Zhao, 2013. "Mining Opinion Words and Opinion Targets in a Two-Stage Framework," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1764–1773.
- Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua, 2011. "Aspect ranking: identifying important product aspects from online consumer reviews," pp. 1496–1505.
- Zhongwu Zhai, Bing Liu, Hua Xu, and Peifa Jia, 2010. "Grouping product features using semi-supervised learning with soft-constraints," pp. 1272–1280.
- 乾孝司・板谷悠人・山本幹雄・新里圭司・平手勇宇・山田薫, 2013 年. 「意見集約における相対的特徴を考慮した評価視点の構造化」, 『自然言語処理』, Vol. 20, No. 1, pp.3–25.
- 小林のぞみ・乾健太郎・松本裕治・立石裕治・福島俊一, 2005 年. 「意見抽出のための評価表現の収集」, 『自然言語処理』, Vol. 12, No. 3, pp.203–222.