

レビュー集合要約における特徴的評価視点セレクション Selecting characteristic attributes for customer review summarization

板谷 悠人†
Yuto Itaya

乾 孝司†
Takashi Inui

山本 幹雄†
Mikio Yamamoto

1. はじめに

近年の Web の発展は目覚ましいものがあり、ホテル予約サイトや、オークションサイトといった各種のサービスが Web 上に提供されている。また、Web サービスのユーザ数の増加に伴って、Blog や掲示板、レビューサイトのように個人が気軽に自分の意見・感想を書き込めるような場も増えてきている。

Web 上に溢れるこれら個人の意見・感想はユーザが商品を買う際の目安になるなど、様々な利用シーンが想定されるが、これらのデータは大量にあり、また、必ずしも適切な構造化がなされているわけではないため、人手により一つ一つこれらの意見・感想を読んでいくには莫大な時間がかかってしまう。この問題を解決する一つの方法として、そういった大量のデータがテキスト解析を通じて、コンパクトにまとめ上げた要約情報を提示することが考えられる。商品に対するレビューサイトを例にすると、ユーザがある対象(商品)についての評価をする際に評価をするポイント(以下、評価視点と呼ぶ)とその評価極性(肯定的評価または否定的評価)を基にして要約を提示すればよい。そうすることによって、ユーザの見やすい形で要約を提示でき、多くのレビューを読むことなく、瞬時に評判を知ることができる。

図1の点線の左側は Zhai ら[2] が示した要約例である。我々もこのようなグラフを要約として出力することを目指す。これはある2つのホテルのレビュー群をそれぞれ一つのグラフに要約したものである。横軸は評価視点情報を示し、「サービス」、「立地」など、予め分類クラスを決めておき、抽出された評価視点をそのどれかに分類する。縦軸はそれぞれのクラスに対して肯定的な意見を述べられた数、否定的な意見を述べられた数を示す。このようなシステムにより、例えばホテルAの「サービス」という評価視点に対して、肯定的な意見が否定的な意見より多いことなどを瞬時に把握することができる。また、ホテル同士を比べることでホテル間の違いが見て取れる。言語処理技術を活用することで、大量のデータ群から図1のような要約情報を自動生成、提示する技術を開発することができる。

レビュー集合から評価視点を網羅的に抽出する課題を Hu ら[1]、板谷[3] では扱っている。本稿はこれらの手法によって抽出された評価視点のうち、どの評価視点をユーザに提示するかというセレクションの部分に焦点を当てる。抽出した全ての評価視点を出力するには数が多すぎるため、提示すべき評価視点を選択する必要がある。評価視点の中には、その対象特有の視点があり、重要な特徴を示している可能性が高い。そこで今回、特徴的評価視点を選び出すということを論文の目的とする。

図1はホテルA、ホテルBに対してそれぞれ我々のシス

† 筑波大学大学院 システム情報工学研究科 コンピュータサイエンス専攻

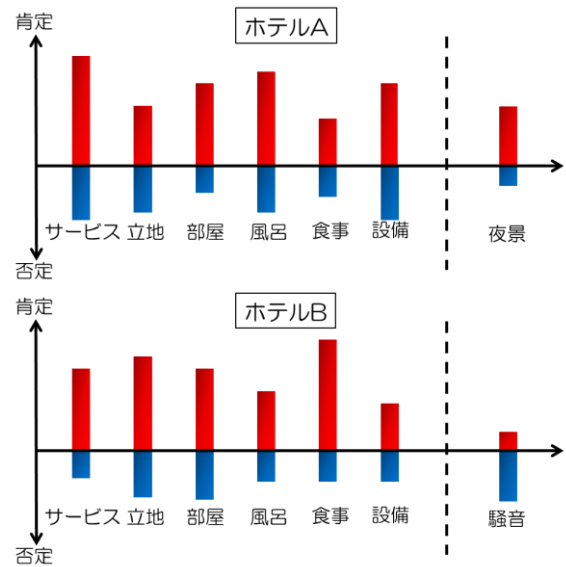


図1 ユーザレビューの要約例

テムを用いて要約を行ったイメージであるが、図の左側はどのホテルのレビューにも現れる共通の評価視点なのに対して、右側はそのホテルの特徴を示している評価視点である。図1の左側は例えば、Zhai らの手法で実現できる。本研究では特に図1の右側を自動作成する手法を新たに提案する。

本論文の構成は以下の通りである。まず、第2節にこの研究の関連研究について述べる。第3節では我々が目的としている手法の全体像について述べる。第4節では始めに特徴的評価視点とは何であるかについて述べ、その次に提案手法について述べる。第5節では評価実験について述べ、第7節では本研究のまとめについて述べる。

2. 関連研究

評価視点をもとに要約することを目的とした研究として Hu らがある。Hu らは文書から評価視点と極性の抽出を行った。評価視点を抽出する方法として、相関ルールを用いている。Hu らは評価視点を頻出評価視点と非頻出評価視点に分け、まず相関ルールを用いて効率よく頻出評価視点を抽出した。次に頻出評価視点の近くにある形容詞は評価表現だと仮定し、評価表現を集めた。最後に頻出評価視点の含まれない文から、集めた評価表現の近くにある名詞フレーズを非頻出評価視点であるとして抽出した。

Hu らの方法で抽出した評価視点の中には、同じ意味を示すにもかかわらず、違う語またはフレーズで表現されるものも多くある。そこで、Zhai らは評価視点のグルーピングを行った。グルーピング方法として、半教師付き学習に緩い制約を用いた手法を提案した。あらかじめ、人手により評価視点がどのクラスに分類されるかというラベル付き

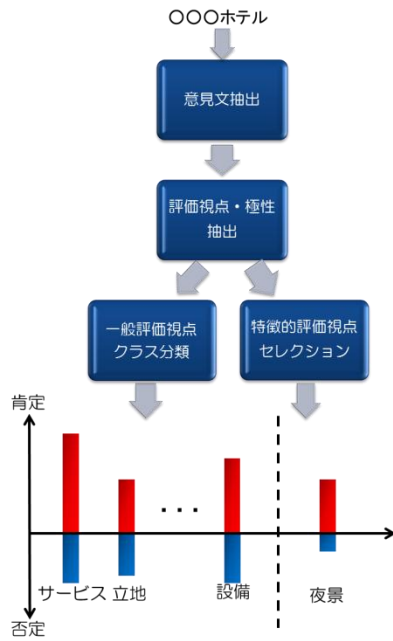


図2 処理の流れ

データを用意しておき、それを少数の学習データとして残りを分類した。分類の素性として評価視点の周辺単語を用いた。Zhaiらの手法ではあらかじめクラス分類すべき評価視点クラスを決めておくため、どのクラスにも当てはまらない新たな評価視点も、あらかじめ定義されたいずれかのクラスへ分類されてしまう。また、彼らは評価視点クラスが、ある商品に固有で、特徴的な視点となりうるかどうかは考慮していない。

我々の研究目的にもっとも類似したものにYuら[7]がある。Yuらは重要な評価視点は多くの人に述べられ、また、商品全体の評価に大きく影響を与えると考え、これら2つをもとに重要度にもとづく評価視点のランキングを行った。Yuらの手法ではユーザが商品进行评估する重要な視点を選び出すのに対して、我々の手法は他の商品と比較するための、商品独自の評価視点を選び出すため目的が異なる。

3. 全体の概要

図2は我々の要約手法全体の処理の流れである。例えば、あるホテルについての評価を知りたいとき、ユーザは「〇〇ホテル」というクエリを入力する。すると、意見文抽出フェーズにおいて、Webから「〇〇ホテル」についての意見が書かれている文書を探してくる。Web上ではユーザが自由に意見・感想を書き込むことができる。

評価視点・極性抽出フェーズでは探してきた意見・感想から「立地」や「食事」といった評価視点を抽出する。また、その評価視点がどういった評価をされているか、その評価極性を解析する。

一般評価視点クラス分類フェーズでは、あらかじめいくつかのクラスを作成しておき、評価視点・極性フェーズにおいて抽出されたものの中から、各評価視点がどのクラスに属するかを分類する。また、特徴的評価視点抽出フェーズでは一般的評価視点クラス分類フェーズと並行して評価視点の中から類似する他の商品と比較してその対象商品

(例の場合は「〇〇〇ホテル」)に特徴的な評価視点を選び出力する。

4. 特徴的評価視点セレクション

4.1 特徴的評価視点

我々は特徴的評価視点をデータ集合中で個別性の高い情報であると定義する。個別性として、ホテル間で比較した場合とレビュー間で比較した場合の2つを考えると、評価視点は表1の4つに整理・分類することが出来る。

表1 評価視点の整理・分類

		ホテル間の個別性	
		なし	あり
レビュー間の個別性	なし	1	2
	あり	3	4

我々はこの中で、ホテル間の個別性が高いものである2および4の自動抽出を研究の目的とする。さらに、2は個人の嗜好性に依存しない特徴的な評価視点であり、4は個人の嗜好性に依存した特徴的評価視点である。我々はこのうち、2を当面のターゲットとする。4は特定のユーザしか述べておらず、全てのユーザに提示するのは良くない。4に該当する評価視点の利用シーンとしては、例えば、個人のプロフィールを利用するようなケースにおいて、同じ嗜好性を持ったユーザに提示するといったことは有効であると考えられる。

4.2 セレクション

数多くある評価視点の中から、評価対象にとって特徴的な視点を選び出すために、評価視点をランキングすることを考える。評価視点の特徴度を表す尺度を定義し、この尺度に基づいて各評価視点にスコアを割り振り、スコアに従ってランキングを行う。提示段階では、この中から最大スコアをもつ評価視点を選び出せば良いことになる。

上記尺度として、我々は対象データにおける評価視点の個別性(出現偏り)を捉えるために、以下で示すような対数尤度比に基づく尺度を定義した。これは内山ら[4]が提案した特定の分野における単語の特徴度を測る尺度を特徴的評価視点セレクションに応用したものである。

4.3 提案手法

以下ではホテルのレビュー集合を要約する例を通して、提案手法を説明する。まず、評価対象およびその比較評価対象となるホテルのレビュー集合中に出現したn個の評価視点を要素とする集合を $V = \{v_1, v_2, \dots, v_n\}$ とする。また、V中に含まれる各評価視点タイプの集合を $U = \{u_1, u_2, \dots, u_m\}$ とする。

例えば、 $V = \{\text{朝食, 価格, 風呂, 朝食, 風呂}\}$ であった場合、その評価視点タイプを表す集合は $U = \{\text{朝食, 価格, 風呂}\}$ となる。

次に、ホテル h_k における各評価視点タイプ u_j の特徴度を測定するために、評価視点 v についての2つの確率変数 W, T を以下のように定義する。

$$W = \begin{cases} 1 & v \text{ の評価視点タイプが } u_j \text{ である。} \\ 0 & \text{それ以外。} \end{cases} \quad (1)$$

$$T = \begin{cases} 1 & v \text{がホテル} h_k \text{のレビューに出現。} \\ 0 & \text{それ以外。} \end{cases} \quad (2)$$

すると、それぞれの変数の関係は表2のようになる。

表2 それぞれ変数の関係

	$T=1$	$T=0$
$W=1$	a	b
$W=0$	c	d

次に以下のような仮説 H_{indep}, H_{dep} を定義する。

- H_{indep} : 確率変数 W と T とは互いに独立である。
- H_{dep} : 確率変数 W と T とは互いに依存している。

上記2つの仮説のもとで、次のような対数尤度比(Log-LikelihoodRatio、LLR)を考える。以下の計算をデータ中の全ホテルの全ての評価視点タイプ u_j について行うことで、個々のスコアを見積もり、各ホテルごとに u_j のランキングを行うことで、特徴度の高い評価視点を選び出す。

$$\begin{aligned} \text{LLR}(u_j, h_k) &= \log \frac{\Pr(V; H_{dep})}{\Pr(V; H_{indep})} \\ &= \sum_{i=1}^n \log \frac{\Pr(W = w_i, T = t_i; H_{dep})}{\Pr(W = w_i, T = t_i; H_{indep})} \\ &= a \log \frac{\Pr(W = 1, T = 1; H_{dep})}{\Pr(W = 1, T = 1; H_{indep})} \\ &\quad + b \log \frac{\Pr(W = 1, T = 0; H_{dep})}{\Pr(W = 1, T = 0; H_{indep})} \\ &\quad + c \log \frac{\Pr(W = 0, T = 1; H_{dep})}{\Pr(W = 0, T = 1; H_{indep})} \\ &\quad + d \log \frac{\Pr(W = 0, T = 0; H_{dep})}{\Pr(W = 0, T = 0; H_{indep})} \\ &= a \log \frac{an}{(a+b)(a+c)} + b \log \frac{bn}{(a+b)(b+d)} \\ &\quad + c \log \frac{cn}{(c+d)(a+c)} + d \log \frac{dn}{(c+d)(b+d)} \quad (3) \end{aligned}$$

この対数尤度比は、2変数 W と T とが依存しているという条件、および、独立であるという条件の下で、データ V が観測される確率の比の対数である。したがって、対数尤度比が大きいときは、2変数は互いに強い依存関係を持つことを意味する。つまり、ホテル h_k において、評価視点タイプ u_j の出現確率が特に高い、または低いことを意味し、この時、評価視点タイプ u_j がホテル h_k にとって特徴的であると考えられる。

この尺度は依存性の度合いを測定するものである。そのため、この尺度では、ある評価視点が特定のホテルのレビューにおいて高確率で出現しているのか、低確率で出現しているのかを区別していない。そこで特定のホテルのレビューにおいて高確率で出現するものに正の値、低確率で出現するものに負の値をとるように式を変形する。

$$\text{LLR}'(u_j, h_k) = \text{sign}(ad - bc)\text{LLR}(u_j, h_k) \quad (4)$$

ただし、

$$\text{sign}(z) = \begin{cases} +1 & z > 0 \\ -1 & \text{それ以外} \end{cases} \quad (5)$$

とする。

4.1節で述べたが、今回は表1の2に該当する、個人の嗜好に依存しない特徴的な評価視点を抽出する。したがって、ユーザ間の意見の違いは考慮しないので、レビューという単位を無視して、各ホテルのレビュー集合をあたかも巨大な一つの文書であるかのように扱う。逆に4の個人の嗜好に依存した特徴的な評価視点を抽出したいときにはレビューというテキスト単位を考慮することは重要である。

4.4 スムージング

前節の尺度は実用上、1つ問題がある。式(3)において、 a, b, c, d のどれか1つでも値が0になってしまうと計算不能になってしまう。例えば、あるホテルには出てくるが、他のホテルには全く出てこないような評価視点は $b = 0$ になってしまう。そこで、スムージング(加算法)[6]を適用する。加算法では、全評価視点の出現頻度に微小の値を追加することで、表2の各セルが0になることを回避する。

5. 実験

5.1 概要

評価視点のランキングを行い、その精度を評価する。評価視点の抽出方法を入れ替えることによって2種類の実験を行う。実験1では特徴的な評価視点のセレクションのみの評価をする。入力として、人手によって得られた正しい評価視点集合を与え、そのランキングの精度を測る。実験2ではシステムを総合的に評価するために、文献[3]で作成した評価視点抽出と特徴的な評価視点セレクションを逐次的に実行し、評価を行う。

5.2 データセット

どちらの実験においても、本実験では楽天トラベル[5]のレビューデータから、東京にあるホテル11件、各100レビューずつを用いて実験を行う。この技術の使用場面を想定すると、ユーザは自分の目的地付近のホテル同士を比べたいはずである。そこでデータセットを東京にあるホテルと地域を限定した。これらのレビューから人手または評価視点抽出器[4]によって評価視点を抽出した。また、その中で特徴的な評価視点とそれ以外とを人手で区別している。人手によって抽出した評価視点を用いて、特徴的な評価視点セレクション単体の評価を行い(実験1)、評価視点抽出器によって抽出した評価視点を用いて、一連の処理全体の評価を行う。(実験2)

実験に用いるデータの詳細を表3に示す。実験1で用いるデータは人手で作成しているため、全て確かな評価視点である。実験2で用いるデータは抽出器が自動で抽出しているため、データの中には評価視点でないものや、レビュー中には出現したものの抽出器によって抽出されなかったものもある。そのため、正解数が実験1に比べて少なくなっている。

5.3 評価尺度

今回評価にはMean Average Precision(MAP)を用いて評価を行う。MAPは情報検索における、ランキングを評価する際によく用いられる尺度であり、式(6)で定義される。

表3 評価視点詳細(1ホテルあたり平均)

	データ	数
人手 (実験1)	評価視点	101.45
	正解数	10.18
評価視点抽出器 (実験2)	評価視点	68.18
	正解数	4.45

$$\text{MAP}(Q) = \frac{1}{|Q|} \sum_{k=1}^{|Q|} \frac{1}{m_k} \sum_{j=1}^{m_k} \text{Precision}_{kj} \quad (6)$$

Q はホテルの集合、 m_k はホテル h_k における特徴的評価視点のタイプ数、 Precision_{kj} はホテル h_k におけるランキングのトップから j 番目の評価視点が出てくるまでの適合率である。

また、比較するベースラインとして tf 法、tf-idf 法を用いる。tf 法は各評価視点が出た数を自身のスコアとする。通常の tf-idf 法はある文書における単語の特徴度を測るものである。今回は文書という単位を無視して、あるホテルにおける評価視点の特徴度を測りたいため、同じホテルのレビュー集合を1つの文書として値を求める。

5.4 実験結果

実験結果を表4に示す。提案手法において、対数尤度比にもとづく尺度を LLR、さらにスムージングを行った結果を LLR+s とした。実験1、実験2どちらにおいても、ベースラインの tf 法および、tf-idf 法に比べ、提案手法(LLR+s)が高い精度を示した。tf 法では比較すべき他ホテルの評価視点情報を全く考慮していないため、精度が低いと考えられる。tf-idf 法では、idf を用いることで、他ホテルについての評価視点情報を考慮することができるが、他ホテルにおける出現の有無の情報しか考慮できず、その出現頻度の情報は無視される。一方、提案手法では他ホテルにおける出現頻度の情報も考慮でき、データ中の評価視点に関する情報を総合的に活用できるようになり、その結果良好な性能が得られたと考えられる。また、実験結果からスムージングが有効であったことが分かる。これは、低出現頻度の評価視点に対しても正確な計算ができるようになったため、性能が向上したと考えられる。

表5に、あるホテルにおける実際のランキング結果上位10件を示した。表中の下線が引かれている評価視点が今回特徴的であるとしたものである。これは実験1の結果である。評価視点の右にある数字は抽出された評価視点の tf 値である。結果から提案手法はベースラインに比べて、出現回数の低い特徴的評価視点もランキング上位へランクさせることが可能である。

6. おわりに

我々はレビュー集合を要約することを目的に研究を行っている。本論文ではその中でも特徴的評価視点のセレクションについて報告した。特徴的評価視点セレクションの手法として、対数尤度比に基づく尺度を提案し、ランキング問題として扱った。その結果ベースラインの手法に比べ、性能が高いことを示した。

特徴的評価視点セレクションでは、ユーザが同じものを指している文字列として異なっていれば、違う評価視点として扱われてしまう。例えば、表5の評価視点の中で、

表4 セレクション実験結果

手法	実験1	実験2
tf	0.418	0.307
tf-idf	0.526	0.309
LLR	0.471	0.317
LLR+s	0.615	0.377

表5 ランキング結果上位10件

ランク	提案手法		tf-idf 法	
	評価視点	tf 値	評価視点	tf 値
1位	<u>パン</u>	11	朝食のパン	10
2位	<u>朝食のパン</u>	10	部屋	29
3位	<u>武道館</u>	6	対応	25
4位	館内	4	<u>パン</u>	11
5位	<u>対応</u>	25	<u>武道館</u>	6
6位	<u>フロント</u>	7	駅	14
7位	<u>朝食</u>	10	館内	4
8位	立地条件	7	立地	10
9位	<u>朝食サービス</u>	3	コンビニ	10
10位	ホテル	10	ホテル	10

「パン」と「朝食のパン」や「朝食」と「朝食サービス」などは同じものについて述べられている。粒度によっては全て同じであるとも考えることもできる。表現数が多いと、それだけ述べられる回数がばらけてしまい、正確にスコア付けすることが出来ないということが起こり得る。このような同じ意味を持つ評価視点をまとめ上げることで、ランキングの性能が向上すると考えられる。

謝辞

本研究は、楽天データチャレンジの際に楽天技術研究所から提供して頂いたデータを用いて行いました。関係者各位に心から感謝いたします。

参考文献

- [1] Minqing Hu, Bing Liu. Mining Opinion Features in Customer Reviews. Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-2004), pp.755-760, 2004.
- [2] Zhongwu Zhai, Bing Liu, Hua Xu, Peifa Jia. Grouping Product Features Using Semi-Supervised Learning with Soft-Constraints*. Proceedings of the 23rd International Conference on Computational Linguistics, pp.1272-1280, 2010.
- [3] 板谷悠人. 共起情報を用いたレビュー文書からの評価視点抽出. 平成21年度筑波大学第三学群情報学類卒業研究論文, 2010.
- [4] 内山将夫, 中條清美, 山本英子, 井佐原均. 英語教育のための分野特徴単語の選定尺度の比較. 自然言語処理, Vol.11, No.3, pp165-197, 2004.
- [5] 楽天トラベル <http://travel.rakuten.co.jp/>
- [6] 辻井純一, 北研二. 確率的言語モデル, 東京大学出版会, 1999.
- [7] Jianxing Yu, Zheng-Jun Zha, Meng Wang, Tet-Seng Chua. Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. pp.1496-1505. 2011.