JAIST Repository
https://dspace.jaist.ac.jp/

Title	レビ　ューテキストの書き手の評価視点に対する評価点の推定
Author(s)	張，博
Citation	
Issue Date	2017-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/14154
Rights	
Description	Supervisor:白井　清昭，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

レビューテキストの書き手の評価視点に対する評価点の推定

北陸先端科学技術大学院大学
情報科学研究科

張 博

平成 29 年 3 月

修 士 論 文

レビューテキストの書き手の評価視点に対する評価点の推定

1510034　　張 博

指導教官　白井清昭

審査委員主査　白井清昭
審査委員　池田心
審査委員　長谷川忍

北陸先端科学技術大学院大学
情報科学研究科

平成 29 年 2 月

概 要

近年，レビューサイトなど，ユーザが製品またはサービスに対するレビューを投稿し，また評価点をつけることができるウェブサイトが増えている．ユーザレビューは消費者とメーカーの両方に有用な情報を含む．消費者は商品の購入やサービスを利用するとき，評価点を参照し，ユーザレビューを読むことにより，各々の好みに合う商品やサービスの選択ができる．また，メーカーはユーザレビューを参照しながら，既存商品あるいはサービスの改善ができるし，新しい商品やサービスの開発もできる．

しかし，膨大の量のレビューテキストを読むのは負担が大きい．そのため，レビューテキストから，ユーザが評価対象の評価視点（例えば，携帯の画面，バッテリの持続時間，カメラなど）に対して与える評価点を予測できれば，ユーザの大まかな評価を手早く知ることができる．本論文では，与えられたレビューテキストから，評価視点毎にユーザの評価点を推定する手法について述べる．

提案手法は大きく 2 つのステップから構成される．ステップ 1 では，レビューテキストから各評価視点に対するコメント (文) を抽出する．提案手法では，評価視点を示唆するキーワードの有無をチェックし，そのキーワードを含むコメントを抽出する．キーワードは，評価視点そのものを表わす単語，全てのレビューにおける出現頻度が上位 100 件の名詞から選別されるもの，既存のシソーラスに登録されるもの及び並列表現を利用して得られるものを使用する．

ステップ 2 では，評価点を推測する．評価視点ごとに，L2 正則化ロジスティック回帰を用いて，評価点 (1〜5 の整数) を推測モデルを学習する．各モデルの入力として，ステップ 1 から抽出されたコメントのみを使う．機械学習の素性として，自立語，評価語またはこれらの否定表現の組み合わせを使用する．

本研究の評価実験について述べる．実験データを作成するために，楽天トラベルからホテルのレビューをクロールする．楽天トラベルのホテルレビューでは，ユーザは 6 つの評価視点（サービス，立地，部屋，設備・アメニティ，風呂，食事）に対して 1 から 5 までの評価点をつけることができる．クロールしたホテルレビューの内，6 つ全ての評価視点の評価点が与えられたレビューのみを訓練及びテストデータとして使用する．

ステップ 1 の評価視点に対するコメントの抽出手法を評価する．訓練データからランダムに選択した 500 件のレビューを対象として，コメント抽出の再現率，精度及び F 値を人手で算出した．「部屋」以外の評価視点について，精度は 90%を越えた．一方，再現率は，「設備・アメニティ」以外の評価視点について 70%以上であった．「設備・アメニティ」の再現率は高くなかったが，設備やアメニティを表わす単語は多岐にわたり，提案手法ではその全てをキーワードとして用意できなかったためと考えられる．

次に，ステップ 2 の評価点の推定手法を評価する．提案手法を 2 つのベースラインと比較する．ベースライン 1 は，評価視点に関するコメントを抽出せず，レビューの全文を用いて評価点を推測する手法である．ベースライン 2 は，ベースライン 1 と同様にレビュー

の全文を用いるが，提案手法と同じ数のレビューを訓練データとして用いる手法である．正解率と平均二乗誤差 (RMSE) を指標として，5 分割交差検定により，各評価視点の評価点推測モデルを評価する．6 つの評価視点の全てを対象としたときの提案手法の正解率は 0.5208 であり，ベースラインより 0.0178 高かった．RMSE は,「立地」「食事」「風呂」の評価視点については，提案手法はベースラインを上回ったが,「サービス」「部屋」「設備・アメニティ」については下回った．一方，正解率については,「設備・アメニティ」以外の評価視点について，提案手法はベースラインを上回った．さらに，推測された評価値と真の評価値の差が非常に大きい事例を 100 件人手で調べ，誤りの要因を分析した．その結果，評価視点と関係のない誤ったコメントが抽出されたこと，評価語が評価視点とは異なる対象に対する意見を表わしていること，などの原因があることがわかった．今後の課題として，コメント抽出方法の改善や，係り受け解析を利用して評価語の対象を正確に判定することなどが挙げられる．

目 次

第 1 章　序論 **1**
- 1.1　研究の背景 1
- 1.2　研究の目的 2
- 1.3　本論文の構成 3

第 2 章　関連研究 **4**
- 2.1　商品属性語抽出に関する研究 4
- 2.2　レビュー文の極性判定に関する研究 6
- 2.3　評価視点からのレビューの分析に関する研究 7
- 2.4　本研究の特色 9

第 3 章　提案手法 **10**
- 3.1　概要 10
- 3.2　評価視点に対するコメントの抽出 11
  - 3.2.1　評価視点を示唆するキーワードのリストの作成 11
  - 3.2.2　評価視点に関するコメントの抽出 12
- 3.3　評価点の推測 15

第 4 章　評価実験 **19**
- 4.1　実験データ 19
- 4.2　実験設定 22
- 4.3　実験結果 23
  - 4.3.1　評価視点を示唆するキーワードのリストの作成 23
  - 4.3.2　評価視点に関するコメント抽出の評価 27
  - 4.3.3　評価点推測の評価 29
- 4.4　誤り分析 31

第 5 章　結論 **39**
- 5.1　まとめ 39
- 5.2　今後の課題 39

謝辞 **41**

第1章　序論

1.1　研究の背景

近年，オンラインショッピングサイトやレビューサイトなど，ユーザが製品またサービスに対するレビューを投稿し，また評価点をつけることができるウェブサイトが増えている．ユーザレビューには消費者とメーカーの両方にとって有用な情報を含む．消費者は，商品を購入したりサービスを利用するとき，ユーザレビューを読み，ユーザによって与えられた評価点を参照することで，各々の好みに合う商品あるいはサービスの選択ができる．一方，メーカーは，ユーザレビューを活用することで，既存の商品やサービスの改善，新しい商品やサービスの開発，販売戦略の見直しなどが行える．また，富士通総研の調査により，直近1回のPCネットショッピングの際に参考にした情報のうち,「購入者の商品レビュー」が最も使われる回数が多いことが報告されている[1]．

しかし，膨大の量のレビューテキストを読むことの負担は大きい．そのため，評判情報解析あるいはオピニオンマイニングと呼ばれる技術が盛んに研究されている．評判情報解析とは，製品やサービスといった特定の対象に対するユーザのコメントを分析し，その対象の評判を明らかにする技術である．例えば，ユーザのレビューが評価対象に対して肯定的な意見を述べているか，それとも否定的な意見を述べているのかを自動的に判定する．この処理は極性判定と呼ばれる．また，意見が表明しているのは評価対象のどの属性なのかを同定した上で極性判定を行う研究も行われている．さらに，大量のユーザレビューを肯定的な意見と否定的な意見に分類し，整理した上で，ユーザに提示するシステムも研究されている．消費者や企業は，大量のレビューを逐一読む代わりに，評判情報解析の結果を参照することで，評価対象に対する評判を簡単に知ることができる．

評判情報分析の結果をユーザに提示する方法のひとつとして，ユーザの対象に対する評価を点数で表わした評価点を示すことが挙げられる．より正確には，レビューテキストから，ユーザが評価対象の評価視点に対して与える評価点を予測する．評価視点は，評価対象となる製品やサービスの属性に対応する．例えば，評価対象が携帯電話なら，画面，バッテリの持続時間，カメラなどの評価視点がある．与えられたユーザレビューに対し，様々な評価視点から見た評価点を自動的に推定し，提示すれば，ユーザの大まかな評価を手早く知ることができる．例えば，ある評価視点を重視する消費者は，全ての商品の中からその評価視点の評価点が高い商品を中心に調べることで，より早く自分の好みに合う商

[1]インターネットショッピング2010 調査書
http://www.fujitsu.com/jp/group/fri/report/cyber/report/shopping2010.html

品を選択できる．一方，企業にとっても，ユーザの評価視点に対する評価点を知ることができれば，多くのレビューテキストを読む必要はないため，自社の商品やサービスの長所や短所を発見する時間が短縮できる．例えば，「製品 A のバッテリ持続時間は高い評価点を得ていることが多い」「製品 B のカメラは高い評価と低い評価が分かれている」といった分析が可能になる．

1.2 研究の目的

本研究では，与えられたレビューテキストから，評価視点毎にユーザの評価点を推定することを目的とする．評価視点は商品の属性やサービスの項目などに対応する．一般に，レビューテキストには，対象に対する全ての評価視点に関するコメントが含まれているわけではなく，ユーザが言及していない評価視点も存在する．例を図 1.1 に挙げる．

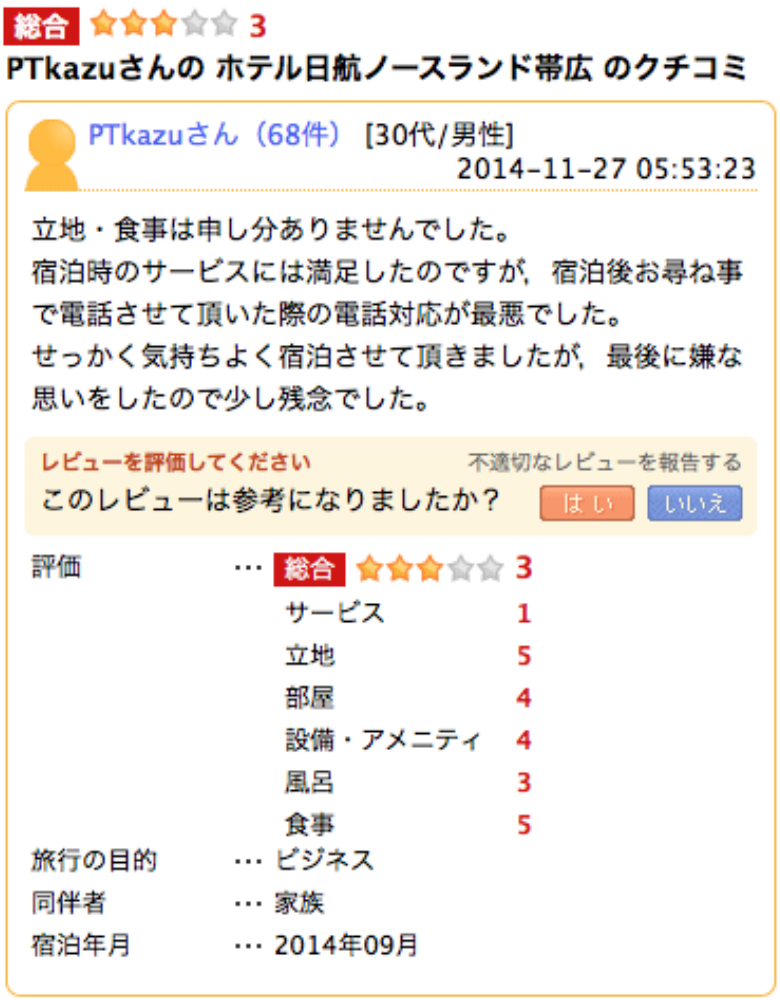


図 1.1: レビューテキストの例

この図のレビューでは,「サービス」という評価視点に関する意見は述べられているが,「部屋」という評価視点については述べられていない．このようなコメントに対し,「部屋」の評価点を推測することは不可能である．本研究では，まずユーザレビューの中に評価視点に関する意見があるかを判定し，意見が存在するときのみに評価点の推測を行う点に特徴がある．また，ユーザレビューの中には，ある評価視点については肯定的だが，別の評価視点については否定的なものもある．このとき，ユーザレビューには肯定的な単語と否定的な単語が混在していると考えられるため，ユーザレビューの全文を用いると，個々の評価視点に対する評価点の推測が難しくなる．そのため，本研究では，評価点を推定する際，ユーザレビューの全文を用いるのではなく，評価視点に言及した文を抽出し，その文のみを手がかりとする．

ユーザが製品やサービスのレビューを投稿できるサイトの中には，ユーザが評価点をつけることができるものもある．しかし，評価点が明示されていないレビューも多数存在する．そのようなユーザレビューに対し，提案手法により評価点を推測することで，より多くのレビューに対してユーザが与えた評価点を集約することが可能になる．

1.3　本論文の構成

本論の構成を以下に述べる．第 2 章では，先行研究について述べ，先行研究と本研究の違いについて論じる．第 3 章では，レビューテキストから評価視点に対する評価点を推測する手法について説明する．第 4 章では，提案手法の評価実験について述べ，実験結果を考察する．また，評価点の予測の誤りの原因を分析する．第 5 章では，本研究のまとめと今後の課題について述べる．

第2章　関連研究

本章では，本研究の関連研究について述べる．2.1 節では，レビューテキストから商品の属性語を自動抽出する研究を紹介する．2.2 節では，レビューテキストの極性判定の研究を紹介する．2.3 節では，評価視点からレビューを分析する手法を説明する．2.4 節では，本研究と関連研究の違いについて論じる．

2.1　商品属性語抽出に関する研究

既に述べたように，本論文では，評価視点に対する評価点を推測する際，評価視点に言及した文を抽出し，それを手がかりに評価点を推測する．評価対象が商品や製品のとき，評価視点は製品の属性に対応することが多い．例えば，製品がノート PC のとき,「CPU」「OS」「バッテリ持続時間」「サイズ」などが属性となる．一方，評判情報分析では，ユーザの評価の対象を明確にするために，製品の属性を表わす語 (属性語) を抽出する研究が行われている．この節では属性語の自動抽出に関する先行研究を紹介する．

Yu らは，商品のレビューテキストから商品の重要な属性を自動的に抽出する手法を提案した [4]．彼らは，先ず賛否意見を含むレビューテキストから出現頻度の高い名詞句を抽出した．次に，クラスタリングを利用し，表記は異なるが同じ属性を指す語，すなわち異表記の属性語をひとつに統合した．最後に，ガウス分布に基づく機械学習アルゴリズムを利用し，商品の属性を重要度の高い順にランキングした．

駒田らは，商品を評価しているツイートから商品の属性を表わす語を自動抽出し，属性語辞書を構築する手法を提案した [7]．Twitter では 140 文字までの投稿制限があるため，短文の投稿が多い．多様な表現がある一方，文長の制限により，文法的に正しくない文も多く存在する．そのため，Twitter 以外のテキストを対象とした既存の手法をそのまま適用するのは適切ではない．この問題を解決するため，駒田らは，ノイズ除去の手法，およびパタンマッチによって属性語を獲得する手法を新たに提案した．属性語抽出の精度が従来の手法より 30%改善したと報告している．

駒田らは，ツイートから以下に該当するものをノイズとして除去する．

- URL を含むツイートの全体
- 「リプライ」「リツイート」「ハッシュタグ」の部分。
- 特定の符号 :“!”，“?”，“･”，“『』”，“「」”，“【】”，“［］”，“（）”，“。”

- 商品に関わらず，頻出する単語

次に，CaboCha を用いて，ツイートの係り受け解析を行う．また，事前に Wikipedia タイトルリストにより MeCab の辞書を拡張しておく．係り受け解析の結果を参照し，評価語と係り受け関係にあり，かつ以下の品詞または品詞列を持っている単語または単語列を属性語の候補とする．

- 名詞-一般，名詞-固有名詞，名詞-サ変接続，未知語
- 接頭詞-数接続　＋　名詞-数
- 接頭詞-名詞接続　＋　名詞-一般
- 連続出現する　名詞-一般

また，係り受け解析の結果に関わらず，以下のパタンにマッチする形態素列を属性語の候補とする．

- 評価語　＋　名詞-一般／名詞-固有名詞／名詞-サ変接続／未知語
- 名詞-一般／名詞-固有名詞／名詞-サ変接続／未知語　＋　評価語
- 評価語　＋　接頭詞-数接続　＋　名詞-数
- 評価語　＋　接頭詞-名詞接続　＋　名詞-一般
- 名詞-一般／名詞-固有名詞／名詞-サ変接続／未知語　＋　名詞-接尾　＋　評価語
- 接頭詞-数接続　＋　名詞-数　＋　評価語

このように獲得された属性語の候補に対し，式 (2.1) により属性語の関連度を計算し，閾値を越えていれば属性語辞書に追加する．

$$R_{xi} = \frac{f_{xi}}{N_x} \times \frac{c_{xi}}{M_x} \tag{2.1}$$

式 (2.1) の記号の意味は以下の通りである．

x: 対象語
i: 属性語の候補
N_x: 対象語 x を含むツイート集合において，評価語と係り受け関係にある文節の総数
M_x: 対象語 x を含む商品評価ツイート集合において，評価語と係り受け関係にある文節の総数
f_{xi}: 対象語 x を含むツイート集合において，属性語候補 i が出現する頻度
c_{xi}: 対象語 x を含む商品評価ツイート集合において，属性語候補 i と評価語が共起する回数

また，この手法では，評価語と共起頻度の高い単語を属性語として獲得するが，評価語辞書に含まれる評価語が多いほど，より多くの属性語が獲得できる．そのため，属性語の抽出と同じように，評価語の抽出も行い，新たに得られた評価語を評価語辞書に追加する．評価語の候補を抽出するためのパターンは以下の通りである．

- 動詞-自立　＋　形容詞-非自立
- 名詞-ナイ形容詞語幹　＋　助動詞-ナイ
- 名詞-ナイ形容詞語幹　＋　助詞-格助詞-一般　＋　形容詞-自立
- 形容詞-自立　＋　助動詞-ナイ
- 形容詞-自立　＋　助詞-係り助詞　＋　助動詞-ナイ
- 属性語　＋　形容詞／名詞-形容動詞語幹／名詞-ナイ形容詞語幹
- 形容詞／名詞-形容動詞語幹／名詞-ナイ形容詞語幹　＋　属性語

次に，抽出された評価語の候補に対し，これまでに獲得された属性語との関連度を式 (2.1) で計算し，それが閾値よりも高いものを評価語辞書に追加する．属性語と評価語の獲得を交互に繰り返し，属性語辞書と評価語辞書の両方を拡張する．

2.2　レビュー文の極性判定に関する研究

小林らは，ポジティブ・ネガティブ分析機能を備えたユーザレビュー集約管理システムを提案した [5]．この手法の処理の流れを図 2.1 に示す．まず句点などの記号を利用し，レビューテキストを文単位に分割する．次に，MeCab を使い，分割された文の形態素解析を行う．その後,「形容詞」「形容動詞」「感動詞」「副詞」「連体詞」「接続詞」「接頭詞」「接尾詞」「名詞」「動詞」の品詞を持つ単語のみを取り出す．その結果，レビューテキスト R を n 個の形態素の集合で表現する．最後に，機械学習されたナイーブベイズ分類器により，文ごとに「ポジティブ」「ネガテイブ」「どちらにも属さない」のいずれかの評価カテゴリに属する確率を計算し，最大確率のカテゴリに分類する．

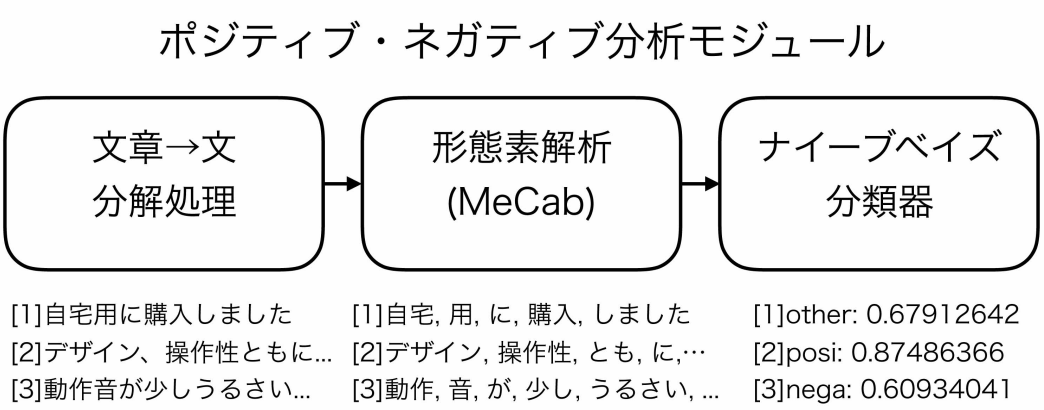


図 2.1: ポジティブ・ネガテイブ分析モジュール

ナイーブベイス分類器では，レビューテキスト R が各カテゴリ C_i に属する確率を式 (2.2) と定義する．

$$P(C_i|R) = \frac{P(R|C_i) \times P(C_i)}{P(R)} \tag{2.2}$$

各記号の意味は以下の通りである．

R: レビューテキスト
C_i: カテゴリ
$P(R|C_i)$: R に含まれる単語がカテゴリ C_i のレビュー中に現れる確率 (尤度)
P(C_i): カテゴリ C_i が選ばれる確率 (事前確率)

式 (2.2) の分母はカテゴリによらず一定なので，計算を省略できる．また，$R = \{\ldots T_j \ldots\}$ の T_j は互いに独立に発生すると仮定する．最終的に，以下の式 (2.3) が最大となるカテゴリを選択する．

$$P(C_i) \prod_{T_j \in R} P(T_j|C_i) \tag{2.3}$$

2.3 評価視点からのレビューの分析に関する研究

小池らは、レビューテキストから評価視点別にレビューテキスト要約のための重要文を抽出する手法を対案した [9]．レビューテキストの評価対象では，予め評価視点が定められることが多い．そのため，評価視点別に要約することによって，レビューを利用するユーザは自分が求めたい内容を容易に把握できる．

外山らは，与えられたレビュー文書集合に対し，文書集合全体と各レビューの分散表現を生成し，各レビューにおける複数の評価視点に対する評価点を同時に推測するニューラルネットワークを学習する手法を提案した [10]．まず Distributed Memory model of Paragraph Vectors (PV-DM) を用いて，各レビューの文書ベクトルとその中の各文のベクトルを学習した．PV-DM の学習の概略を図 2.2 に示す．この図は，単語の分散ベクトルを周辺に出現する単語の類似性に基づき学習すること，また文章の分散ベクトルも同時に学習することを示している．

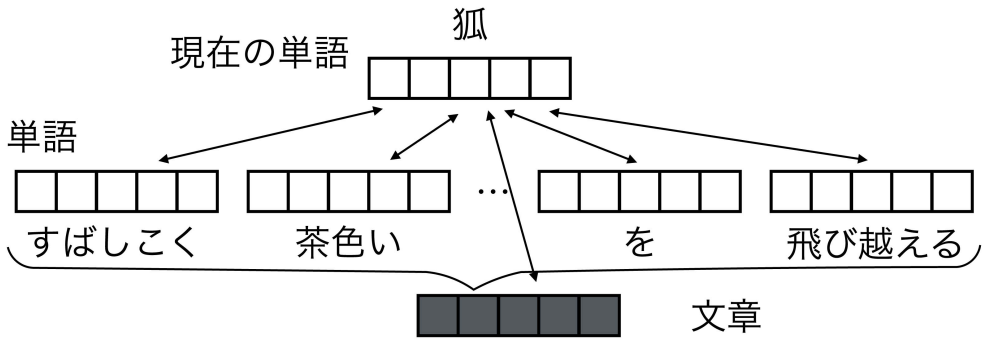


図 2.2: PV-DM の学習の概略 (文献 [10] より)

次に，レビュー内に出現する全ての文ベクトルを，圧縮された文ベクトルに変換する．ここでの目的は，レビューによって文の数が異なるので，全てのレビューに対して同じ数の文ベクトルを入力としてニューラルネットワークの学習を行うことにある．圧縮された文ベクトルは，元の文ベクトルの重み付け平均により求める．ベクトルの重みは，圧縮後の文ベクトルと元の文ベクトルの位置が近いほど高い値を与えるように決める．

最後に，図 2.3 に示すように，全結合ニューラルネットワークから構成された分類器によって，複数の評価視点の評価点を予測する．入力層は圧縮された文ベクトルと文書ベクトルである．出力層は評価視点と評価点の組であり，評価視点の数×評価点のラベルの数だけのユニットを持つ．すなわち，文ベクトルと文書ベクトルで表現されたレビューに対し，各評価視点の評価点のラベル (1〜5 までの整数) を予測するニューラルネットワークを学習する．

実験の結果，評価点推測の正解率は 0.5021 となった．また，提案手法は文ベクトルと文書ベクトルの両方を使う点に特長があるが，文ベクトルのみを用いる手法や文書ベクトルのみを用いる手法に比べて，正解率が 1%〜2%向上したと報告している．

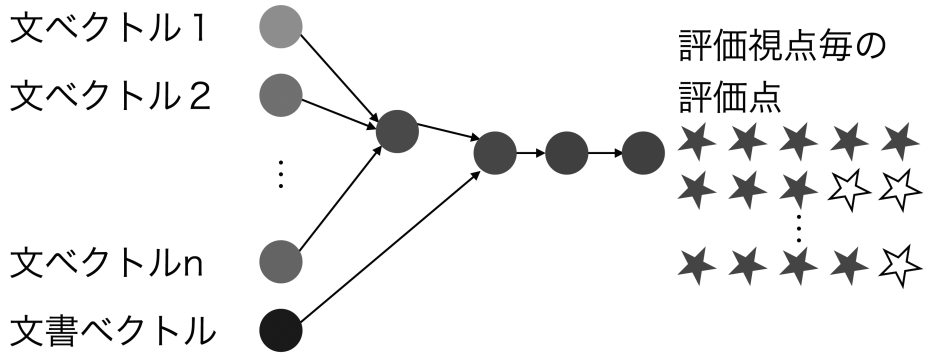


図 2.3: 全結合ニューラルネットワークによる分類器 (文献 [10] より)

Pham らは，least square method を適用し，ホテルのレビューを対象に，評価視点毎に評価点を推測する手法，ならび最も重要な評価視点を推測する手法を提案した [8]．評価点は，レビューの特徴ベクトルと，それと同じ次元を持つ重みベクトルの内積で推定する．重みベクトルは訓練データにおける正解の評価点との誤差が最小になるように学習する．

2.4　本研究の特色

小林らの研究 [5] を始め，評判情報分析はレビューの極性 (肯定的か否定的か) を判定することを目的としたものが多いが，本研究は評価点を推測することを目的とする．また，レビュー全体の評価点を推測するのではなく，評価対象の評価視点毎に評価点を推測する．外山らの研究 [10] や Pham らの研究 [8] でも評価視点毎にユーザの評価点を予測しているが，評価視点に関する言及の有無に関わらず，常に評価点を推測するという問題点がある．評価視点について対象を評価した文が存在しないときには評価点を推測するのは不可能である．本研究ではレビューテキスト内に評価視点に対するコメントが存在するかをまず判定し，存在する場合のみ，レビューに含まれる全ての文ではなく，その評価視点に関するコメントのみを入力として評価点を推測する．

第3章　提案手法

本章では，与えられたレビューテキストに対し，評価視点に対するユーザの評価点を推測する手法を提案する．評価対象はホテルとする．すなわち，ホテルのレビューを入力とし，その評価点を推測する．

3.1　概要

本研究では，評価視点はあらかじめ定義されているものとする．また，評価点を推測するためのデータとして楽天トラベル[1]に投稿されたユーザレビューを用いる．楽天トラベルでは，ユーザは，ホテルに関するコメントを投稿し,「サービス」「立地」「部屋」「設備・アメニティ」「風呂」「食事」および「総合」のそれぞれの観点から評価点を付けることができる．本研究では「総合」を除く6つをホテルの評価視点と定義し，それぞれに対する評価点を推測するモデルを学習する．それぞれの評価視点の説明を以下に記す．

サービス スタッフの客への対応などを評価する．

立地 駅や繁華街に近いかなど，ホテルの場所を評価する．

部屋 広さ，禁煙室か喫煙室かなど，部屋を評価する．

設備・アメニティ テレビ，インターネット接続などの設備や，歯ブラシやシャンプーなどの備品を評価する．

風呂 シャワールームや温泉など，入浴設備を評価する．

食事 朝食や夕食など，ホテルが提供する食事を評価する．

楽天トラベルでは，ユーザが与える評価点は1から5までの整数である．これに倣い，本研究で推測する評価点は1から5までの整数とする．

本研究の提案手法は以下の二つのステップから構成される．1番目のステップでは，評価視点を示唆するキーワードの有無をチェックし，レビューテキストからそのキーワードを含むコメント (文) を抽出する．2番目のステップでは，評価視点毎に，ステップ1で抽出されたコメントに対して評価点を推測するモデルを適用し，1〜5の評価点を出力する．

[1] http://travel.rakuten.co.jp

ただし，ステップ 1 でコメントが抽出できなかった評価視点については，評価点は推測せず,「不明」と判定する．評価点を推測するモデルは，評価視点毎に，楽天トラベルにおけるレビューと評価点を訓練データとして，教師あり機械学習により学習する．図 3.1 は上記の処理を図示したものである．

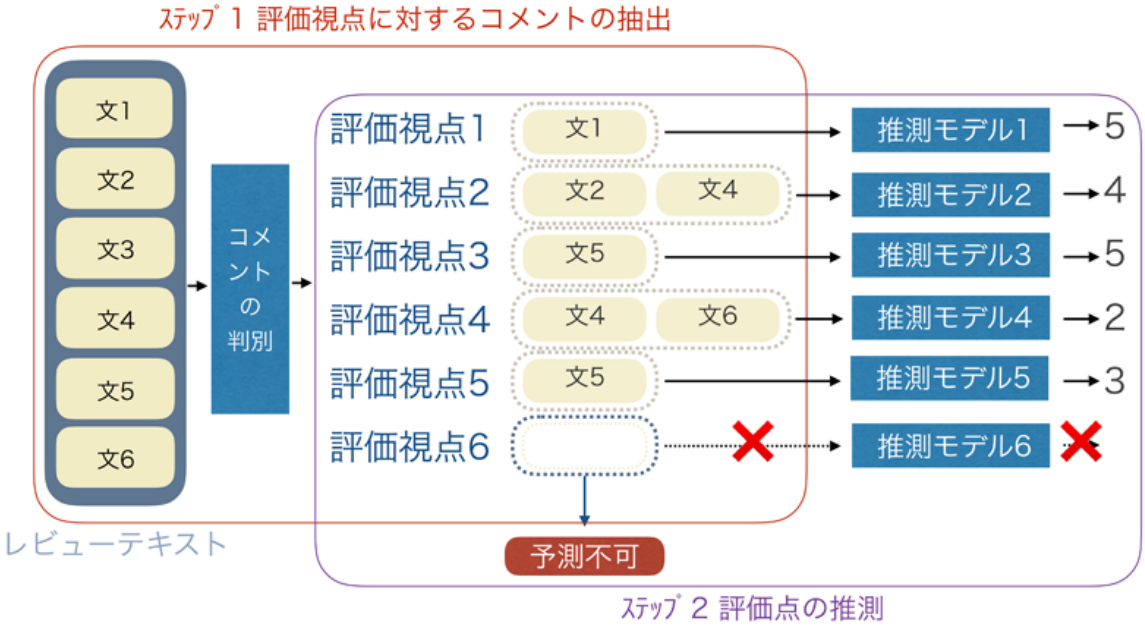


図 3.1: 提案手法の概要

3.2 評価視点に対するコメントの抽出

ここでは，レビューテキストから評価視点に対するコメントを抽出する手法について説明する．まず，各評価視点を示唆するキーワードのリストを作成する．詳細は 3.2.1 項で述べる．次に，作成されたキーワードを用いて，レビューテキストからコメントを抽出する．詳細は 3.2.2 項で述べる．

3.2.1 評価視点を示唆するキーワードのリストの作成

各々の評価視点を示唆するキーワードは以下の手順で作成する．

1. 評価点を表す単語そのものを追加する．

   ・「サービス」「立地」「部屋」「設備」「アメニティー」「風呂」「部屋」をそれぞれの評価視点のキーワードとする．

2. Weblio 類語辞典[2] を参照し，ステップ 1 で得られた単語の類義語を追加する．

   ・例えば,「部屋」の類義語である「ルーム」と「室内」を追加する．

3. 楽天トラベルから収集したレビューテキストのコーパスから，ステップ 1 と 2 で獲得されたキーワードを含むレビューテキストを取得する．この中における出現頻度上位 100 件の名詞から，評価視点のキーワードとして適しているものを人手で選別し追加する．ただし，既にステップ 1 やステップ 2 で得られた単語，ならびにストップワード[3] は除く．

   ・例えば，評価視点「サービス」のキーワードとして,「対応」「フロント」などを追加する．

4. 並列表現を手掛かりとしてキーワードを獲得する.「A (と | や | また | および) B」という並列表現があり，A がキーワードリストにあるとき，B を新たにキーワードとして追加する．

   ・例：レビューテキストの中に「シャンプーとリンス」があり,「シャンプー」が既にキーワードリストに存在するとき,「リンス」を新たなキーワードとして追加する．

   なお，このステップは，評価視点が「設備・アメニティ」のキーワードを獲得するときのみ用いる．ホテルには様々な設備があり，また顧客に提供するアメニティグッズの種類も多いため,「設備・アメニティ」という評価視点のキーワードは多岐にわたると考えられる．ステップ 1 から 3 でキーワードを獲得するだけでは不十分と判断し，並列表現を手がかりとしたキーワードの獲得を試みる．

5. 既存のシソーラスを利用する．具体的には，日本語語彙大系 [3] の「料理 (0847)」というカテゴリに登録されている単語を評価視点「食事」のキーワードに追加する．

   ・例：「パン」「すき焼き」「懐石料理」など

6. 上記の手続きでキーワードが不十分と判断したときは，ステップ 3 の処理を繰り返して，キーワードを追加する．ただし，評価視点「設備・アメニティ」については，ステップ 3 と 4 を繰り返す．

図 3.2 は上記の手続きを示したものである．

3.2.2 評価視点に関するコメントの抽出

ここでは，レビューテキストから評価視点のキーワードを含むコメント (文) をその評価視点を評価する文として抽出する．

[2] http://thesaurus.weblio.jp/

[3] http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt

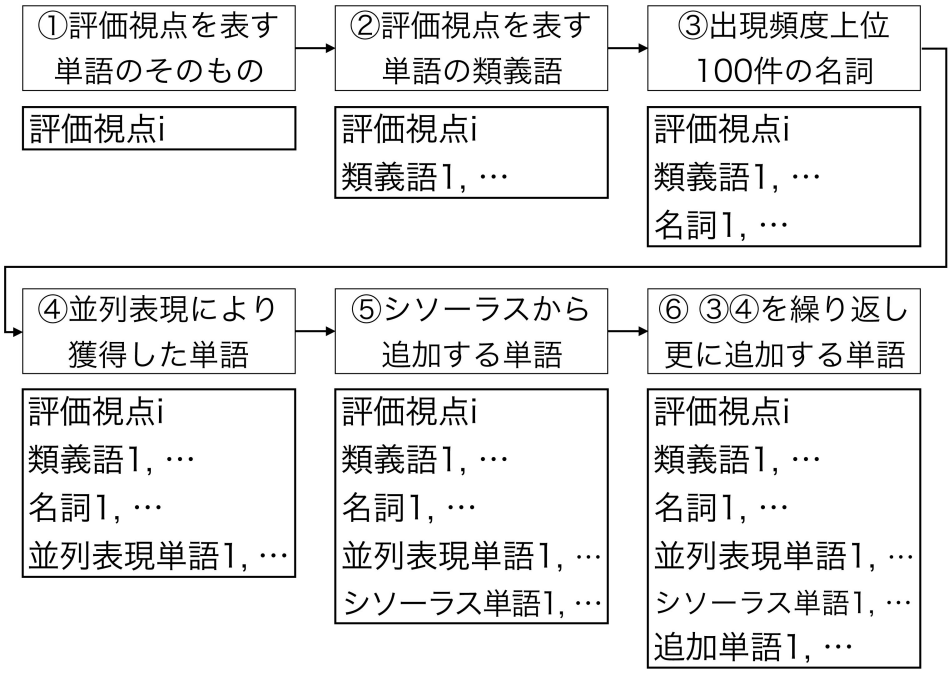


図 3.2: キーワードリスト作成の手順

レビューテキストの文分割

まず，事前にレビューテキストを文単位に分割する．レビューテキストでは句読点は様々な記号で表現される．レビューテキストを適切に分割するために，句読点の表記揺れも考慮する必要がある．そのため，レビューテキストを分割する前に，表 3.1 に示すように記号を変換し，句読点を正規化する．

表 3.1: 句読点の正規化

元の記号	変換した記号
,(半角)　，(全角)	、
.	。
!	！
?	？

そして，表 3.2 に挙げた 3 つの記号を句点とし，これを文末として，レビューテキストを文単位に分割する．

表 3.2: 文分割のための句点

句点	。　？　！

レビューテキストからのコメントの抽出

評価視点のキーワードの中には複数の単語から構成されているものもある．すなわち，単語の 2-gram や 3-gram がキーワードとして登録されていることがある．例えば，評価視点「サービス」のキーワードのひとつに「従業+員」があるが，これは「従業」と「員」の 2 つの単語から構成されている．

単語 n-gram のキーワードとマッチさせるために，まずレビュー文を形態素解析する．形態素解析には MeCab[4] を用いた．次に，あらかじめ用意されたキーワードを含む文を評価視点に関するコメントとして抽出する．単語 n-gram がキーワードのとき，キーワードを構成する単語を順番にマッチさせ，全ての単語がマッチしたときにその文を取り出す．「従業+員」の場合，まず「従業」という単語を探し，その次の単語が「員」のとき，マッチングに成功したとみなす．

コメント抽出の例

レビューテキストから評価視点に関するコメントを抽出する例を図 3.3 に示す．この例では,「立地」という評価視点のコメントを抽出する過程を示す．まず，対象のレビューテキストの記号を正規化し，文単位に分割する．次に，文毎に形態素解析を行い，評価視点毎にキーワードをチェックする．キーワードを含む文をその評価視点の評価文として抽出する．ここで,「JR 三ノ宮駅から本当に近くて。」の中に評価視点「立地」のキーワード「駅」が含まれるため，この文を評価視点「立地」のコメントとして抽出する．

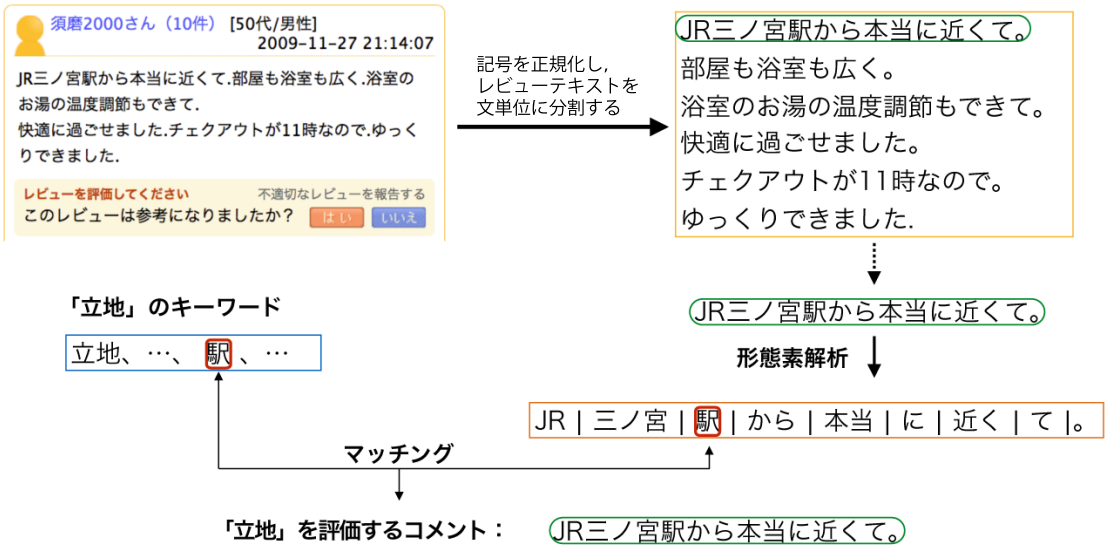


図 3.3: コメント抽出の流れ

[4] http://taku910.github.io/mecab/

3.3 評価点の推測

評価点の推測では，評価視点毎に評価点 (1〜5 の整数) を推測するモデルを L2 正規化ロジスティック回帰によって学習する．学習ツールとして LIBLINEAR[1] を用いる．学習のパラメタはデフォルト値を選択する．

L2 正則化ロジスティック回帰では，データを素性ベクトルで表現する必要がある．本研究では，ステップ 1 で抽出された評価視点に関するコメントを素性ベクトルで表現する．素性ベクトルの作成に用いる素性ならびにベクトルにおける素性の重みを表 3.3 に示す．

表 3.3: 学習素性とその重み

素性タイプ	素性	重み
自立語	⟨単語⟩	0.2
自立語の否定	⟨単語⟩-否定	0.2
評価語	⟨単語⟩-評価語	1.0
評価語の否定	⟨単語⟩-評価語-否定	1.0

以下，それぞれの素性について説明する．

- 自立語

  コメントに含まれる自立語を素性とする．自立語の中には評価点に関連性の高いものが含まれる．例えば,「静かだ」は部屋の評価視点が高いことを,「騒がしい」は部屋の評価視点が低いことを示唆する．また,「とても」「大変」は評価点が 5 点または 1 点に近いことを,「まあまあ」は 3 点に近いことを示唆する．

- 自立語の否定

  コメントに含まれる自立語の次に否定を表わす表現が続くとき,「⟨単語⟩+否定」を素性とする．ただし，⟨単語⟩は自立語の基本形である．自立語に否定表現が続くときは，異なる極性を表わすと考えられる．例えば,「静かだ」は評価点が高いことを示唆するが,「静かじゃない」は評価点が低いことを示唆する．このため，前者の素性は「静かだ」，後者の素性は「静かだ+否定」とすることで両者を区別する．

- 評価語

  評価語とは,「よい」「悪い」など，ある対象の良し悪しを評価する単語である．このような評価語は評価点と強い関連性があると考えられるため，素性として用いる．ただし，自立語と区別するために,「⟨単語⟩+評価語」という形式で素性を表現する．本研究では，日本語評価極性辞書 (用言編 [6] と名詞編 [2]) に登録されている単語を評価語と定義する．

- 評価語の否定

  自立語と同様に，評価語に否定表現が続くときは,「⟨ 単語 ⟩+評価語+否定」を素性とする．ここで⟨ 単語 ⟩は評価語の基本形である．

自立語と評価語では，評価語の方が評価点を推測するための強い手がかりになると考えられる．そのため,「評価語」ならびに「評価語の否定」の素性の重みは 1.0 とし,「自立語」ならびに「自立語の否定」の素性の重みは 0.2 と設定する．これらの重みは直観により定めた．

自立語と評価語の抽出

素性ベクトルを作成するために，コメントから自立語と評価語を抽出する．ステップ 1 から得られたコメントを MeCab によって形態素解析し，以下の品詞を持つ単語を自立語として抽出する．

> 名詞-一般，名詞-サ変接続，名詞-形容動詞語幹，名詞-副詞可能，名詞-未知,
> 動詞-自立，形容詞-自立,
> 副詞-助詞類接続，接頭詞-名詞接続 [5]

一方，既に述べた通り，日本語評価極性辞書の用言編と名詞編に含まれる単語を評価語として抽出する．用言編の辞書には 2311 個の評価語が，名詞編の辞書には 5297 個の評価語が収録されている．日本語評価極性辞書では，評価語が複合語，すなわち単語の n-gram で表わされていることがある．本研究では，$n = 1, 2, 3, 4$ の単語 n-gram で表現される評価語を用いる．

否定の判定

本研究ではパタンマッチにより，否定表現の有無を判定する．否定表現を検出するためのパターンを図 3.4 に示す．同図において，⟨ 単語 ⟩は自立語もしくは評価語を表わす．自立語もしくは評価語に続く単語列が図 3.4 のいずれかにマッチするとき,「⟨ 単語 ⟩+否定」もしくは「⟨ 単語 ⟩+評価語+否定」という素性を抽出する．

素性の抽出例

素性の抽出例を挙げる．図 3.5 のレビューから抽出される素性を表 3.4 に示す表 3.4 は，評価視点毎に対して抽出したコメント，そのコメントから抽出した素性を示している．また，素性が「自立語の否定」「評価語」「評価語の否定」である場合，その素性を太字と下線で表示している．

[5] http://www.nltk.org/book-jp/ch12.html

⟨単語⟩　＋　ない／なく
⟨単語⟩　＋　じゃ／は／で／も／が　＋　ない／なく
⟨単語⟩　＋　できる／する　＋　ない／なく
⟨単語⟩　＋　で＋は／で＋も　＋　ない／なく
⟨単語⟩　＋　て＋いる／する＋てる／できる＋てる　＋　ない／なく
⟨単語⟩　＋　ませ＋ん (⟨単語⟩が「ある」「する」「できる」以外の場合)
⟨単語⟩　＋　じゃ／は／で／も／が　＋　ある　＋　ませ＋ん
⟨単語⟩　＋　する／できる　＋　ませ＋ん
⟨単語⟩　＋　する＋て／できる＋て　＋　ませ＋ん

図 3.4: 否定表現を検出するパターン

出張に利用。リッチは車での利用なので駅からの距離は関係なし。夜の食事処も徒歩圏内で問題なく、安くておいしいお店も紹介いただきありがたかったです。朝食は十分満足。コストパフォーマンスはとても良いです。スタッフの方の対応もとても気持ちよく満足でした。

図 3.5: レビューテキストの例

表 3.4: 抽出された素性の例

評価視点	抽出したコメント	素性
サービス	スタッフの方の対応もとても気持ちよく満足でした。	スタッフ, 対応, とても, 気持ちよい, 満足-評価語
立地	リッチは車での利用なので駅からの距離は関係なし。	リッチ, 車, 利用, 駅, 距離, 関係-否定
部屋	なし	なし
設備・アメニティー	なし	なし
風呂	なし	なし
食事	夜の食事処も徒歩圏内で問題なく、安くておいしいお店も紹介いただきありがたかったです。朝食は十分満足。	夜, 食事, 処, 徒歩, 圏内, 安い, お, 店, 紹介, いただき, ありがたい, 朝食, 十分, おいしい-評価語, 満足-評価語

作成した訓練データの例

以上の手続きにより，訓練データに含まれるテキストから，評価点を推測するための訓練データとなる素性ベクトルの集合を作成する．作成した訓練データの一部を図 3.6 に示す．最初の数字は評価点を表している．その後は「⟨素性のインデックス⟩-⟨素性⟩:⟨素性の重み⟩」から構成される素性ベクトルである．

図 3.6: 素性ベクトルの例

4 7-フロント:0.2 16-部屋:0.2 47-対応:0.2 48-とても:0.2 59-雰囲気:0.2 60-よい:0.2 61-綺麗:0.2 19910-良い-評価語:1.0
5 8-スタッフ:0.2 48-とても:0.2 50-お:0.2 90-掃除:0.2 91-人:0.2 92-含める:0.2 93-ターミナル:0.2 94-感じ:0.2 95-リラックス:0.2 19910-良い-評価語:1.0
3 47-対応:0.2 80-従業:0.2 107-駐車:0.2 111-方々:0.2 114-到着:0.2 161-値段:0.2 167-安い:0.2 168-最初:0.2 169-あまり:0.2 170-期待-否定:0.2 171-びっくり:0.2 172-広い:0.2 173-感動:0.2 19910-良い-評価語:1.0
4 106-接客:0.2 112-態度:0.2 592-挨拶:0.2 1011-ウェイトレス:0.2 19910-良い-評価語:1.0 19935-気持ち良い-評価語:1.0
1 4-サービス:0.2 7-フロント:0.2 25-少し:0.2 34-する:0.2 47-対応:0.2 84-提供:0.2 157-夕食:0.2 234-想定:0.2 385-うれしい:0.2 522-待たす:0.2 523-全体:0.2 19916-満足-評価語:1.0
5 4-サービス:0.2 27-ある:0.2 108-無料:0.2 153-大:0.2 196-思う:0.2 339-浴場:0.2 341-マッサージ:0.2 342-チェア:0.2 524-使える:0.2
.....

第4章　評価実験

本章では，提案手法の評価実験について述べる．4.1 節では実験で用いたデータについて述べる．4.2 節では実験の設定について説明する．最後に，4.3 節で実験の結果について述べる．

4.1　実験データ

本実験では，楽天トラベルに掲載されているホテルのレビューを使用する．楽天トラベルのウェブサイトから，合計 411,568 件のホテルレビューを収集した．ただし，ユーザはコメントを投稿する際に評価点の入力を省略できるので，全てのレビューで 6 つの評価視点の全てに対する評価点が付与されているわけではない．各評価視点の評価点が付与されたレビューの件数を表 4.1 に示す．「サービス」「立地」「部屋」の 3 つの評価視点については評価点を付けることが必須のため，これらのレビュー件数は総件数と同じである．実験では，6 つの評価視点全てに評価点が付与された 272,665 件のレビューを使用した．

表 4.1: 評価視点が付与されたレビューの件数

評価視点	レビュー件数
サービス	411,568
立地	411,568
部屋	411,568
設備・アメニティ	409,132
風呂	392,656
食事	279,432

データ取得の概要

ユーザレビューの取得方法について述べる．本実験では，Python の urllib3[1] と beautifulsoup4[2] の二つのパッケージを利用した．urllib3 は，ウェブサーバ上にあるファイルを

[1]http://urllib3.readthedocs.io/en/latest/index.html
[2]https://www.crummy.com/software/BeautifulSoup/bs4/doc/

HTTP 経由で保存するライブラリであり，ホテルのレビュー詳細ページ (html ファイル) を保存する際に用いた．beautifulsoup4 は，HTML ファイルや XML ファイルから情報を抽出するためのライブラリであり，保存した html ファイルからタグを除いて，レビューテキスト，ユーザ情報，評価点などの情報を抽出する際に用いた．

ホテルのレビューページの URL の取得

楽天トラベルのホテルレビューページ (お客様の声) の URL は以下のように構成される．

> http://review.travel.rakuten.co.jp/hotel/voice/**11327**/?f_time= &f_keyword= &f_age=0&f_sex=0&f_mem1=0&f_mem2=0&f_mem3=0 &f_mem4=0&f_mem 5=0&f_teikei=&f_static=1&f_point=0&f_sort=0&f_next=**0**

図 4.1: 楽天トラベルのホテルのレビューページの例

上記の「**11327**」はホテルのインデックスであり，この数字を変えることで，他のホテルのレビューページの URL も取得できる．最後の「**0**」は，このページで表示する最初のレビューの番号である．この数字は 20 の倍数であることが分かった．楽天トラベルで

は 1 つのページあたり 20 件のユーザレビューを表示する仕様になっているためと考えられる．ホテルのレビューページの例を図 4.1 に示す．このページでは,「お客様の声」の右の括弧にレビューの総数が書いてある．この数字から，上記の URL の“ f_next ”の属性値を 0 からいくつまで変化させれば全レビューを取得できるかがわかる．このように，ホテルのインデックスや“ f_next ”の値を変えながらページをダウンロードすることで，ホテルのレビューを網羅的に収集できる．

レビューテキスト，評価点の取得

図 4.1 のレビューページでは，レビュー内容の一部しか表示されていない．ユーザが書いたレビューの全文は，各レビューのタイトル (図 4.1 で「レビューのタイトル」とマークされているところ) のリンク先のページにある．このページを「レビュー詳細のページ」と呼ぶ．図 4.2 はその例である．レビュー詳細のページには，ユーザ ID，レビューの全文，および 6 つの評価視点ならびに「総合」の評価点が含まれる．

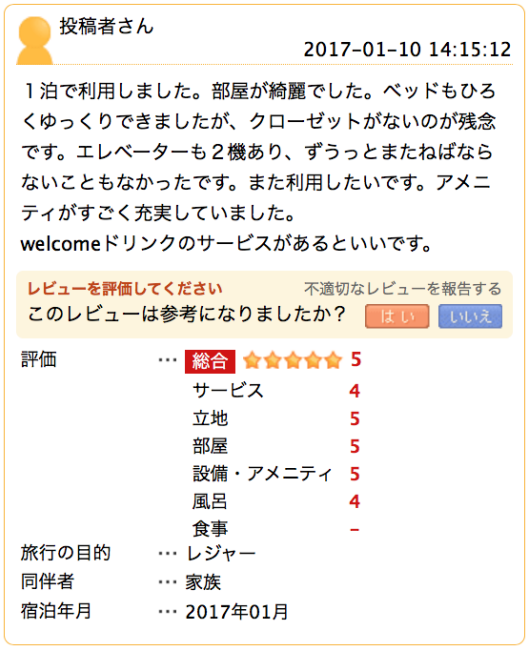
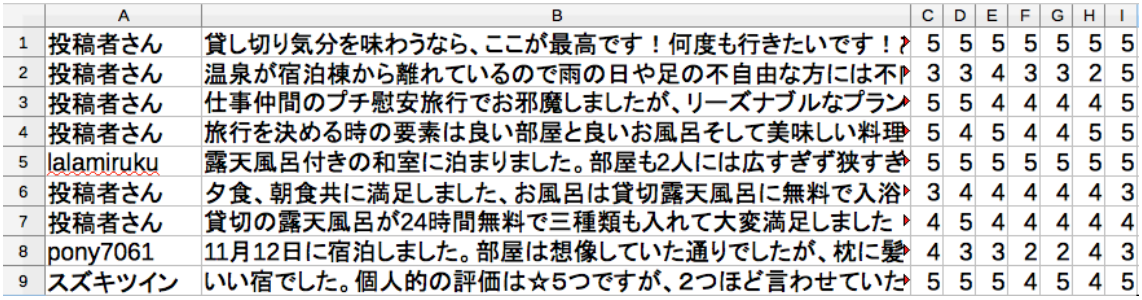


図 4.2: レビュー詳細のページの例

urllib3 を使い，レビューのタイトルのリンク先のページを保存する．次に，保存したページから beautifulsoup4 を使い，レビューテキスト，評価視点の評価点などの情報をホテル毎に取得し，csv ファイルとして保存する．csv ファイルの一行は以下の内容から構成される．

> ユーザ名, レビューテキスト, 総合評価点, サービスの評価点, 立地の評価点, 部屋の評価点, 設備・アメニティの評価点, 風呂の評価点, 食事の評価点

作成した cvs ファイルのサンプルを図 4.3 に示す．



	A	B	C	D	E	F	G	H	I
1	投稿者さん	貸し切り気分を味わうなら、ここが最高です！何度も行きたいです！	5	5	5	5	5	5	5
2	投稿者さん	温泉が宿泊棟から離れているので雨の日や足の不自由な方には不	3	3	4	3	3	2	5
3	投稿者さん	仕事仲間のプチ慰安旅行でお邪魔しましたが、リーズナブルなプラン	5	5	4	4	4	4	5
4	投稿者さん	旅行を決める時の要素は良い部屋と良いお風呂そして美味しい料理	5	4	5	4	4	5	5
5	lalamiruku	露天風呂付きの和室に泊まりました。部屋も2人には広すぎず狭すぎ	5	5	5	5	5	5	5
6	投稿者さん	夕食、朝食共に満足しました、お風呂は貸切露天風呂に無料で入浴	3	4	4	4	4	4	3
7	投稿者さん	貸切の露天風呂が24時間無料で三種類も入れて大変満足しました	4	5	4	4	4	4	4
8	pony7061	11月12日に宿泊しました。部屋は想像していた通りでしたが、枕に髪	4	3	3	2	2	4	3
9	スズキツイン	いい宿でした。個人的の評価は☆5つですが、2つほど言わせていた	5	5	5	4	5	4	5

図 4.3: ホテルレビューの csv ファイル (一部)

4.2 実験設定

本実験では，評価点を推測するシステムの性能を評価する．評価基準は正解率と平均二乗誤差 (Root Menu Square Error;RMSE) とする．正解率は，システムが推測した評価点が正解の評価点と一致している割合である．その定義を式 (4.1, 4.2) に示す．

$$A_i = \frac{R_i}{N_i} \tag{4.1}$$

$$A = \frac{\sum_{i=1}^{6} R_i}{\sum_{i=1}^{6} N_i} \tag{4.2}$$

A_i は評価視点 i の正解率，A は 6 つの評価視点の正解率のマイクロ平均である．R_i は評価視点 i について予測した評価点が正解となったレビューの件数，N_i は評価視点 i について評価点を推測したレビューの総数である．RMSE は，システムが推測した評価点と正解の評価点の誤差を測る指標である．その定義を式 (4.3, 4.4) に示す．

$$RMSE_i = \sqrt{\frac{1}{N_i}\sum_{n=1}^{N_i}(y_n - \hat{y}_n)^2} \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{4.4}$$

$RMSE_i$ は評価視点 i についての平均二乗誤差，$RMSE$ は 6 つの評価視点全ての平均二乗誤差である．y_n は n 番目のレビューについてシステムが推測した評価点，$\hat{y}_n$ は n 番目のレビューの正解の評価点である．N_i は式 (4.1) と同じく評価視点 i について評価点を推測したレビューの総数である．N は 6 つの評価視点について評価点を推測したレビューの総数である．$RMSE_i$ と $RMSE$ は値が小さいほど正解との誤差が小さく，良い結果であることを示す．

実験では以下の 3 つのシステムを比較する．

1. ベースライン 1
   評価視点に関するコメントを抽出せず，レビューに含まれる全ての文を用いて評価点を推測する手法．モデルを学習する際も，コメントの全文から素性を抽出し，素性ベクトルを作成し，これを訓練データとする．提案手法と異なり，評価視点に関する意見がレビュー文に含まれていない場合でも評価点の推測を試みる．

2. ベースライン 2
   ベースライン 1 と同様にレビューの全文を用いて評価点を推測するが，提案手法と同じ量の訓練データを用いて学習されたモデルを用いる手法．提案手法では，レビューテキストから評価視点に対するコメントが存在する場合にのみ，それらのコメントを訓練データとするため，訓練データの量はベースライン 1 より少ない．一般的に，機械学習では訓練データの量が多いほど，正解率が高くなる．評価視点に関するコメントが抽出されたときのみ評価視点を推測すること，また抽出されたコメントのみを評価点の予測に使うことの効果をより厳密に調べるため，訓練データの量が同じという条件でベースラインと提案手法を比較する．ベースライン 2 では，各評価視点に対し，全ての訓練データからランダムで提案手法と同じ量のデータを選択し，ベースライン 1 と同じように全文から素性ベクトルを作成し，評価点の推測モデルを学習する．

3. 提案手法
   提案手法では，レビューテキストの中で各評価視点に対するコメントが抽出できる場合のみ，抽出されたコメント素性ベクトルを作成し，その評価視点の評価点の推測モデルを学習する．

ベースラインと提案手法の違いを図 4.4 と図 4.5 を用いて説明する．図 4.5 は図 3.1 の再掲である．ベースライン 1 とベースライン 2 では，レビューテキストの全文を用いて評価点を推測する．また，6 つの評価視点全てについて常に評価点を推測する．一方，提案手法では，評価視点に関するコメントをまず抽出し，そのコメントのみを用いて評価視点の評価点を推測する．また，コメントが抽出できなかったコメントについては評価点を推測しない．

4.3 実験結果

4.3.1 評価視点を示唆するキーワードのリストの作成

本項では，3.2.1 項で説明した手法で作成した評価視点に関するキーワードについて報告する．キーワードは評価視点に関するコメントを抽出するために用いる．キーワードリストを作成する手順を簡潔に再掲する。

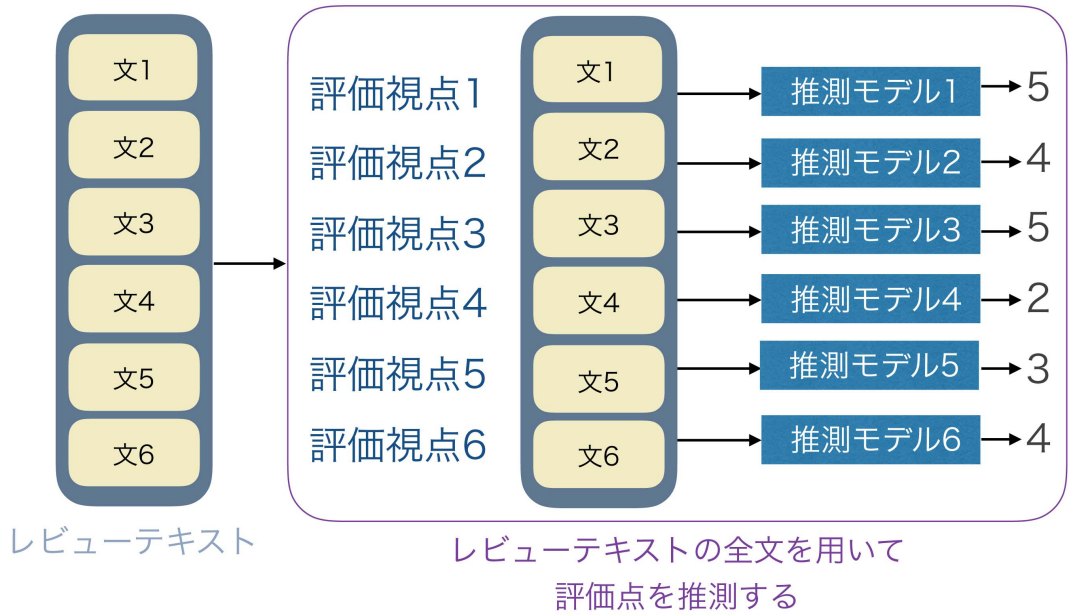


図 4.4: ベースライン手法の概要

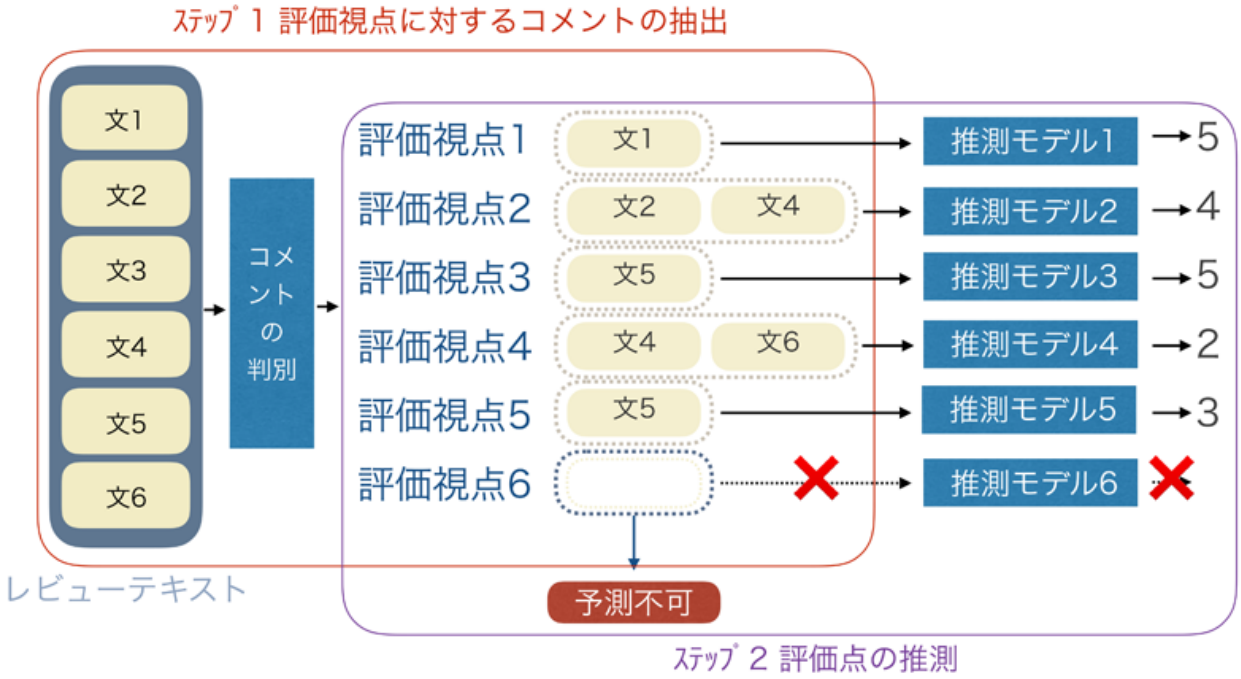


図 4.5: 提案手法の概要 (再掲)

1. 評価視点を表す単語のそのものを追加する
2. Weblio 類似語辞書に掲載される類義語を追加する
3. 出現頻度上位 100 件の単語から人手でふさわしい単語を追加する
4. 並列表現を手がかりに新しいキーワードを取得し追加する (評価視点が「設備・アメニティ」のときのみ)
5. 既存のシソーラスを用いて料理名を表わすキーワードを追加する (評価視点が「食事」のときのみ)
6. ステップ 3 と 4 を繰り返す

各評価視点を示唆するキーワードを取得する過程を表 4.2 から表 4.7 に示す.「ステップ」はキーワードを獲得した手法を表わし，上記の手順の番号に対応している．また，一つのステップが 2 回以上実行される場合，その実行回数も書いてある.「キーワード」は該当ステップで取得したキーワードである.

表 4.2: 評価視点「サービス」のキーワードを取得する過程

ステップ	キーワード
1	サービス
2	(なし)
3(1 回目)	対応, フロント, スタッフ, 従業＋員, 接客, 荷物
3(2 回目)	応対, 受付, 受け付け, うけつけ, 担当＋の, 係員, 係＋の, 係り＋の
3(3 回目)	女将, 宿＋の＋方, 施設＋の＋方

表 4.3: 評価視点「立地」のキーワードを取得する過程

ステップ	キーワード
1	立地
2	場所, ロケーション
3(1 回目)	駅, コンビニ, 利便
3(2 回目)	アクセス, 交通, タクシー
3(3 回目)	JR
3(4 回目)	位置, 繁華街
3(5 回目)	地下鉄, 新幹線
3(6 回目)	街＋から, 近く＋便利, 近くで＋便利

表 4.4: 評価視点「部屋」のキーワードを取得する過程

ステップ	キーワード
1	部屋
2	ルーム, 室内
3(1 回目)	(なし)
3(2 回目)	(なし)
3(3 回目)	客室

表 4.5: 評価視点「設備・アメニティ」のキーワードを取得する過程

ステップ	キーワード
1	アメニティー：設備, アメニティー, アメニティ
2	(なし)
3(1 回目)	ベット
4(1 回目)	布団, 枕
3(2 回目)	タオル, 冷蔵庫, エアコン, エレベーター
4(2 回目)	浴衣, ドライヤ, ドライヤー, スリッパ, インターネット, LAN, ソープ, 暖房
3(3 回目)	ネット, コンセント, シャンプー, シャンプ, リンス, 空調, 便座
4(3 回目)	洗顔, Wi-Fi

表 4.6: 評価視点「風呂」のキーワードを取得する過程

ステップ	キーワード
1	風呂, ふろ
2	バス, 浴室, 浴槽
3(1 回目)	温泉, 露天風呂, 浴場, シャワー
3(2 回目)	入浴, お湯, 男湯, 女湯
3(3 回目)	(なし)
3(4 回目)	源泉

表 4.7: 評価視点「食事」のキーワードを取得する過程

ステップ	キーワード
1	食事
2	ご飯, 御飯, ごはん, 夕飯, 夕食, 晩食, 昼飯, 昼食, ランチ, ディナー
3(1 回目)	朝食, バイキング, レストラン, メニュー
3(2 回目)	朝モーニング
3(3 回目)	朝飯, 料理
5	日本語語彙大系における料理名

最後に，各評価視点のコメントを抽出する際に用いたキーワードの一覧を表 4.8 に示す．

表 4.8: コメント抽出で用いたキーワード一覧

サービス	サービス, 接客, フロント, 受付, 対応, スタッフ, 従業+員, 荷物, 女将, 応対, うけつけ, 受け付け, 係員, 係+の, 係り+の, 担当+の, 宿+の+方, 施設+の+方
立地	立地, 位置, 駅, 繁華街, コンビニ, 距離, アクセス, 交通, 利便, JR, 新幹線, 地下鉄, タクシー, 場所, ロケーション, 市街地, 街+から, 近く+便利, 近く+で+便利
部屋	部屋, ルーム, 室内, 客室
設備・アメニティ	設備, アメニティ, アメニティー, 空調, エアコン, ベッド, 冷蔵庫, エレベーター, インターネット, Wi-Fi, LAN, コンセント, 便座, ウォシュレット, 暖房, 布団, 枕, シャンプー, リンス, ソープ, 洗浄機, 洗浄器, タオル, ドライヤ, ドライヤー, スリッパ, 洗顔, 浴衣
風呂	風呂, ふろ, シャワー, バス, 温泉, 浴室, 浴場, お湯, 男湯, 女湯, 入浴, 露天風呂, 浴槽, 源泉
食事	食事, ご飯, 御飯, ごはん, 料理, 朝モーニング, 朝食, 昼食, 晩食, 夕食, レストラン, メニュー, バイキング, 朝飯, 昼飯, 夕飯, ディナー, おかず, 【日本語語彙大系における料理名】

4.3.2 評価視点に関するコメント抽出の評価

本項では，4.3.1 項で示したキーワードを用いて評価視点に関するコメントを抽出する手法を評価する．実験データから 500 件のレビューをランダムに選択し，6 つの評価視点毎に，その評価視点からホテルを評価しているコメントを人手で付与した．これを正解と

して，実際に自動抽出されたコメントと比較し，コメント抽出手法の精度，再現率，F 値及び抽出率を測った．それぞれの定義を式 4.5 から式 4.8 に示す．

$$\text{精度} : Precision = \frac{\text{抽出した正しいコメントの数}}{\text{抽出したコメントの数}} \tag{4.5}$$

$$\text{再現率} : Recall = \frac{\text{抽出した正しいコメントの数}}{\text{抽出すべきコメントの数}} \tag{4.6}$$

$$\text{F 値} : F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.7}$$

$$\text{抽出率} : Rate = \frac{\text{コメントを抽出できたレビューの数}}{\text{レビューの総数}} \tag{4.8}$$

コメント抽出手法の精度，再現率，F 値の結果を表 4.9 に示す．コメントを抽出できたレビューの件数および抽出率を表 4.10 に示す．

表 4.9: コメント抽出手法の評価

評価視点	精度	再現率	F 値
サービス	90.20	77.67	83.46
立地	94.34	73.10	82.39
部屋	85.17	87.45	86.29
設備・アメニティ	92.42	61.68	73.99
風呂	93.48	75.41	83.48
食事	98.19	81.23	88.91

表 4.10: コメント抽出の抽出率

評価視点	件数	抽出率
サービス	101,725	37.3%
立地	80,679	29.6%
部屋	125,783	46.1%
設備・アメニティ	56,203	20.6%
風呂	100,502	36.9%
食事	166,017	60.9%

コメント抽出の精度は，評価視点「部屋」では 85%であるものの，他の 5 つの評価視点では 90%を越えており，十分に高いことがわかる．しかし，再現率は 61%から 87%と低く，まだ改善の余地がある．原因として，コメント抽出の手法で利用した評価視点のキーワードが不足しているためと考えられる．特に「設備・アメニティ」には，様々なものがあり，これを示唆する関連語の数も多いが，提案手法ではその全てを網羅的に収集できていない．

次に抽出率について考察する．表 4.10 に示した通り，抽出率は 20%から 60%となり，評価視点によって大きく異なることがわかった．また,「設備・アメニティ」の抽出件数が最も少ないが，再現率が低い原因と同様に，コメントを抽出するためのキーワードが不足していることが原因と考えられる．ただし，レビューの中には評価視点に関する意見が書かれていない場合もあるため，全てのレビューから評価文を取り出す必要はない (抽出率を 100%に近づける必要はない) ことに注意していただきたい．

4.3.3 評価点推測の評価

本項では，ユーザの評価点を推測する手法を評価する．本実験では，4.1 節で説明した実験データを用いて，5 分割交差検定により，評価点を推測するモデルの学習及び評価を行った．ベースライン 1，ベースライン 2，提案手法の 3 つの手法について，各評価視点の正解率を図 4.6 と表 4.11 に，RMSE を図 4.7 と表 4.12 に示す．表 4.11 と表 4.12 に示した「平均」は，6 つの評価視点の正解率もしくは RMSE のマイクロ平均である．

表 4.11: 評価点推測の正解率

	サービス	立地	部屋	設備	風呂	食事	平均
BL1	0.5099	0.5021	0.4713	**0.4426**	0.4618	0.4866	0.4791
BL2	0.5066	0.4973	0.4660	0.4297	0.4563	0.4863	0.5030
提案手法	**0.5524**	**0.5688**	**0.4973**	0.4054	**0.5156**	**0.5323**	**0.5208**

表 4.12: 評価点推測の RMSE

	サービス	立地	部屋	設備	風呂	食事	平均
BL1	**0.8750**	0.8875	**0.9306**	**0.9542**	1.0162	0.9677	**0.9398**
BL2	0.8897	0.8968	0.9407	0.9836	1.0207	0.9723	0.9529
提案手法	0.9326	**0.8558**	0.9680	1.0980	**1.0041**	**0.9491**	0.9687

設備 = 設備・アメニティ，BL1 = ベースライン 1，BL2 = ベースライン 2

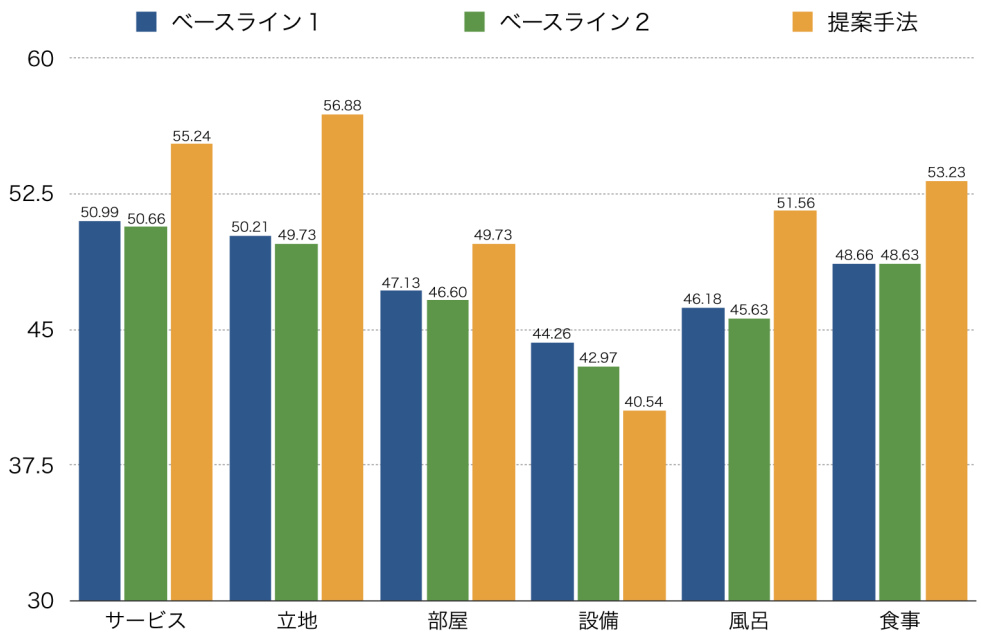


図 4.6: 評価点推測の正解率

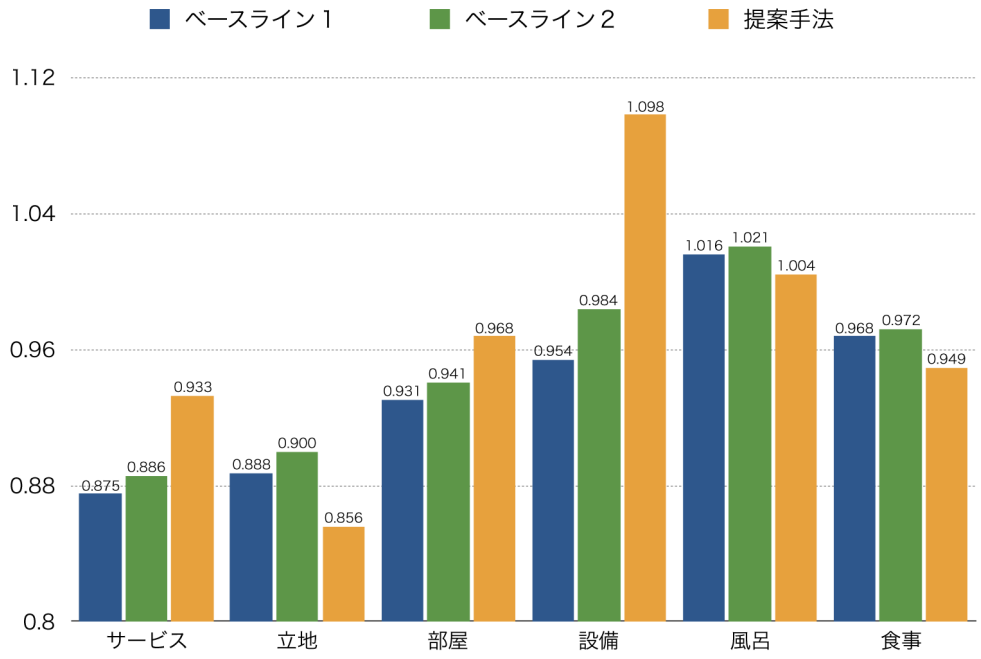


図 4.7: 評価点推測の RMSE

提案手法の正解率は,「設備・アメニティ」を除いた5つの評価視点についてベースラインより高くなった．6つの評価視点の平均でも，提案手法はベースラインよりも正解率が高い．評価視点についてのコメントが存在しないレビューを学習やテストの対象としないことにより，評価点を予測するモデルの性能が向上したことがわかる．

「設備・アメニティ」の正解率がベースラインより低い原因は以下と考えられる．まず，訓練データの量が少ないことが挙げられる．表4.10に示すように，他の評価視点では8万から16万件のレビューを使ったが,「設備・アメニティ」では約5万6千件しか使わなかった．ただし，ベースライン2と提案手法は同じ量の訓練データを用いているが,提案手法の正解率はベースライン2よりも劣る．両者の違いは，提案手法は「設備・アメニティ」に関するコメントのみを訓練やテストに用いているの対し，ベースライン2はレビューの全文を用いているという点である．これは，設備・アメニティに関するコメント以外の文にも,「設備・アメニティ」の評価点の推測に有効な情報が存在することを示唆する．現在の手法では，キーワードを含む文を「設備・アメニティ」の評価点の推測に有用な文として抽出しているが，その抽出方法を見直す必要がある．

なお，6つの評価視点の全てについて，ベースライン1はベースライン2と比べて正解率が高いことから，訓練データの量は正解率に影響を与えることが確認された．

表4.12のRMSEの結果を見ると，6つの評価視点の平均では，提案手法はベースラインを上回る．個々の評価視点毎に比較すると,「立地」「風呂」「食事」の3つの評価視点については，提案手法のRMSEはベースライン1やベースライン2よりも小さい．すなわち，これらの評価視点については，コメントが存在しないレビューを学習やテストの対象としないことにより，評価点を予測するモデルの性能が向上したことがわかる．一方,「サービス」「部屋」「設備・アメニティ」については，ベースラインと比べてRMSEが大きくなった．特に,「サービス」と「部屋」については，正解率とRMSEとで手法の優劣が一致していない．すなわち，提案手法の正解率はベースラインより良い結果が得られた一方，RMSEの結果は悪くなった．つまり,「サービス」と「部屋」については，提案手法はベースラインと比べて，より多くのレビューに対して正解の評価点と同じ点を推測できるが，正解と大きく異なる評価点を推測した場合も多かった．

4.4　誤り分析

前項で報告した通り,「サービス」と「部屋」の評価視点については，提案手法は真の評価点と大きく異なる評価点を推測することが多かった．この原因を調べるための誤り分析を実施する．まず，以下の条件を満たすレビューを誤り分析の対象とする．

- 真の評価点が5で，提案手法が予測した評価点が1であったレビュー
- 真の評価点が1で，提案手法が予測した評価点が5であったレビュー

すなわち，評価点の予測値と真の値の差が大きいレビューを対象とする．条件を満たす「サービス」と「部屋」のレビューの中からそれぞれ50件を選択し，それらのレビューに

対し，推測を誤った要因を人手で分析した．分析に用いた 100 件のレビューのうち，真の評価点が 5 のときに 1 点と推測したレビューの数は 8 件，真の評価点が 1 のときに 5 点と推測したレビューの数は 92 件であり，後者の数が多かった．

誤り分析の結果，誤りを引き起こす要因として表 4.13 に示す 10 種類のものが見つかった．表 4.13 には,「サービス」と「部屋」のそれぞれについて，該当するレビューの件数も示す．

表 4.13: 評価点予測の誤り分析

原因	サービス	部屋
1. 評価文が短く，評価点を推測するための情報が不足している	5	4
2. 評価文が長く複雑で，ノイズとなるような素性が抽出されている	4	8
3. 評価文抽出の誤り	2	5
4. レビューの内容とユーザが付与した評価点が矛盾している	3	0
5. 評価語が辞書に登録されていない	8	3
6. 評価語が別の対象もしくは他の評価視点に対する意見を示している	5	6
7. 否定の判定の誤り	5	5
8. 評価文が肯定的・否定的意見の両方を含む	2	3
9. 形態素解析の誤り	0	1
10. 不明	16	15

次に，それぞれの誤りの原因について，例を挙げながら説明する．

1. 評価文が短く，評価点を推測するための情報が不足している．

   抽出されたコメントが短いほど，評価点を推測するとき使う素性が少なくなる．表 4.14 に例を示す．この表は，評価視点，真の評価点，予測値 (システムが推測した評価点)，レビューの原文，抽出した評価視点に関するコメント，機械学習に用いた素性を示している．抽出したコメントは短く，素性が 4 つしかないため，評価点の推測は難しい．

表 4.14: 評価文が短かいことが原因となった誤りの例

項目	評価視点：部屋　　真の値：1　予測値：5
原文	子連れにはオススメです。なんといってもプールの存在がでかい。部屋は … 本館と新館はまるで別物。本館の和室は凄みがありますぜ。とくにシャンデリア。日本のどこを探しても、こんなのありません。異国情緒が横溢しまくりです。温泉はよかったです。夜は閑散としていたのに朝は激込みでした。馬油シャンプーなどというものがあり、ホテルではイチオシらしい。食事は … 値段を考えればOK。カレーがうちで作る奴よりおいしかった（笑）ただ、北海道らしい料理を期待してはいけません。他所で食べるべし。レストラン会場のテーブルも凄い。テーブルの四脚が木彫の奔馬。グアーっと口をあけて、いなないております。得体の知れない魚のしゃちほこバージョンもあり。探せば、龍や虎もあるに違いない。豪奢この上なし。日本人のショッパイ枯山水趣味なぞブルドーザーで踏みにじって行くかのような古代皇帝趣味。ビューホテル只者ではない。一見の価値あり。
抽出したコメント	部屋は … 本館と新館はまるで別物。
素性	部屋, 本館, 新館, 別

2. 評価文が長く複雑で，ノイズとなるような素性が抽出されている．

   1 の原因とは逆に，抽出されたコメントが長い場合，内容が複雑になったり，ノイズとなるような素性 (判定に悪影響を与えるような素性) が抽出されたりすることが誤りの原因となっているケースである．表 4.15 に示す例では，抽出されたコメントの中に禁煙ルームと喫煙ルームに関する内容が書いてあり，複雑である．一方，次回に宿泊に対する希望について書かれている文があるが，これは部屋の評価とは無関係であり，この文から取り出された素性は判定に悪影響を与えると考えられる．

3. 評価文抽出の誤り．

   本研究で提案するコメント抽出の手法では，常に正確なコメントを抽出できない．このため，評価視点と関連性の低いコメントを誤って抽出する場合がある．抽出されたコメント中の素性を用いて評価点を推測するのは不正確である．表 4.16 に示す例では，コメントに評価視点「部屋」のキーワード「客室」が入っていたため「部屋」のコメントとして抽出された．しかし，このコメントは主に「風呂」を評価するコメントである，この中の素性を用いて「部屋」の評価点を推測はするのは適切ではない．

表 4.15: 評価文が長いことが原因となった誤りの例

項目	評価視点：部屋　　真の値：1　予測値：5
原文	朝食バイキングが素晴らしいです。おなかいっぱいにさせていただきました。今回、喫煙ルームしか空いていなかったので泊まりましたが、是非とも禁煙ルーム化を進めていただきたいと思いました。昨今、必ず「余る側」は喫煙ルームです。今回は窓を開け、換気扇を付けて寝ましたが、季節柄そろそろ、それも限界です。良い宿で、禁煙ルームが空いていたら是非また泊まりたいと思うので、かような希望を書かせていただきました。
抽出したコメント	今回、喫煙ルームしか空いていなかったので泊まりましたが、是非とも禁煙ルーム化を進めていただきたいと思いました。昨今、必ず「余る側」は喫煙ルームです。良い宿で、禁煙ルームが空いていたら是非また泊まりたいと思うので、かような希望を書かせていただきました。
素性	今回, 喫煙, ルーム, 空く, 泊まる, 是非, 禁煙, 進める, 思う, 昨今, 必ず, 余る, 側, 良い-評価語, 宿, また, かよう, 希望, 書く

表 4.16: 評価文抽出の誤りが原因となった誤りの例

項目	評価視点：部屋　　真の値：1　予測値：5
原文	食事が本当にがっかりしました。天ぷらあげてるのが見えるのにできたての天ぷらはたべれず冷えててびっくりしました。バイキング楽しみにしていましたが食堂でたべてるみたいでした。従業員の方も愛想があまり良くないので。このホテルを選んだの失敗だったかなと寂しくなりました。風呂もぬるいし露天風呂は向かいのホテルの客室から見えてるそうです (私の友人が向かいのホテルに泊まった時見えたそうです。)
抽出したコメント	風呂もぬるいし露天風呂は向かいのホテルの客室から見えてるそうです (私の友人が向かいのホテルに泊まった時見えたそうです。)
素性	風呂, ぬるい, 露天風呂, 向かい, ホテル, 客室, 見える, 友人, 泊まる

4. レビューの内容とユーザが付与した評価点が矛盾している．

ユーザが評価点を与えたときに間違えた可能性がある．また，一つの評価視点の評価点を与えるとき，他の評価視点が影響することも考えられる．表 4.17 に示す例を見ると，ユーザは評価視点「サービス」に関して良いコメントしか書いてない．しかし，その評価点は「1」と低い点数である．ユーザが誤って「1」を与えたか，お風呂の掃除に不満を持っていることが評価点に影響した可能性がある．

表 4.17: 矛盾した評価点が原因となった誤りの例

項目	評価視点：サービス　　真の値：1　予測値：5
原文	妹が友人と遊びに来たときに使わせていただきました。立地は抜群です。新宿駅が近いので、どこにいくにも全く不便しません。サービス面も良く、チェックアウト後も荷物を預かっていただいたのはとても助かりました。また、チェックインも、少し早めの到着だったのですが、対応して入れて下さりました。ただし、部屋とお風呂に少し不満がありました。お風呂のお湯の温度が低く、ずっとぬるかったとのこと。またお風呂とベッドに髪の毛があったとのこと。これはかなり申し訳ない気持ちになりました。掃除は徹底して欲しいですし、冬なので湯の温度もしっかりしていただきたいと思います。以降宜しくお願いします。
抽出したコメント	サービス面も良く、チェックアウト後も荷物を預かっていただいたのはとても助かりました。また、チェックインも、少し早めの到着だったのですが、対応して入れて下さりました。
素性	サービス, 良い-評価語, チェック, アウト, 荷物, 預かる, とても, 助かる-評価語, チェックイン, 少し, 早め, 到着, 対応, 入れる, 下さる

5. 評価語が辞書に登録されていない．

一般的に，評価語は通常の自立語より評価点に影響しやすい．そのため，3.3 節で説明した通り，素性ベクトルを作成するときに，評価語の素性の重みを自立語の素性の重みより高く設定している．しかし，評価語辞書に登録されていない評価語がレビュー文に出現していても，それに対して高い重みを与えることができない．表 4.18 に示す例では,「大変」と「不満」の評価語が辞書に載っていなかったため，これらの単語を評価語の素性として抽出することはできなかった．

表 4.18: 評価語が辞書に登録されていないことが原因となった誤りの例

項目	評価視点：サービス　　真の値：1　予測値：5
原文	格安なので仕方がないですが、荷物を自分で運ばなければならなかったので、大変でした。(子供が一緒なので) サービスに不満です。
抽出したコメント	格安なので仕方がないですが、荷物を自分で運ばなければならなかったので、大変でした。(子供が一緒なので) サービスに不満です。
素性	格安, 荷物, 自分, 運ぶ-否定, 大変, 子供, 一緒, サービス, 不満

6. 評価語が別の対象もしくは他の評価視点に対する意見を示している.

   評価語の対象が違うと，正しくない評価語が評価点の推測に使われる．特にその極性が違うとき，評価点の推測に大きな影響を与える．表 4.19 に示す例では,「悲しい」という評価語がある．しかし，この評価語は以前宿泊したときのフロントの対応を評価しており，今回の宿泊のサービスを評価しているわけではない．実際，レビューの原文を読むと，今回のサービスに満足していることがわかる.「悲しい」という評価語は否定的見解を表わすので，評価点は誤って 1 と低く推定されている.

表 4.19: 評価語の対象が違うことが原因となった誤りの例

項目	評価視点：サービス　　真の値：5　予測値：1
原文	いつも利用しています。実は以前のことですが、フロントの方の事ですごく悲しい思いをしたことがありました。今はその様なこともなく、お部屋も快適に利用できているので良いです。バーのお姉さんがいつも気さくで素敵です。たまにしか行かないのに覚えていてくださっていて嬉しいです。
抽出したコメント	実は以前のことですが、フロントの方の事ですごく悲しい思いをしたことがありました。
素性	以前, フロント, 思い, 悲しい-評価語, ある

7. 否定の判定の誤り.

   否定の判定に失敗すると，素性は反対の極性を示唆することになるため，評価点推測への影響が大きい．例を表 4.20 に示す．この例では,「綺麗」という単語は否定されているが，提案手法のパターンマッチングでは否定表現を検出できなかったため，誤った素性が抽出されている.

表 4.20: 否定の判定の誤りが原因となった誤りの例

項目	評価視点：部屋　　真の値：5　予測値：1
原文	綺麗な部屋でなかった。この評価が妥当。風呂は大変狭すぎた。
抽出したコメント	綺麗な部屋でなかった。
素性	綺麗, 部屋

8. 評価文が肯定的・否定的意見の両方を含む.

   コメントが肯定的・否定的意見の両方を含む場合がある．一般的に，逆接接続詞の後ろにある文の方がより強く伝えたい意見を表わしており，その文を評価点を推測に使うべきである．一方，提案手法はそのような処理を行わないため，肯定的な文も否定的な文も素性の抽出に用いていた．表 4.21 に示す例では,「サービス」の評価点を推測するとき，肯定的・否定的両方の意見を使用したため，正確な推測ができなかった.

   表 4.21: 肯定的・否定的意見の両方が含まれることが原因となった誤りの例

項目	評価視点：サービス　　真の値：1　予測値：5
原文	立地・食事は申し分ありませんでした。宿泊時のサービスには 満足したのですが、宿泊後お尋ね事で電話させて頂いた際の電 話対応が最悪でした。せっかく気持ちよく宿泊させて頂きまし たが、最後に嫌な思いをしたので少し残念でした。
抽出したコメント	宿泊時のサービスには満足したのですが、宿泊後お尋ね事で電 話させて頂いた際の電話対応が最悪でした。
素性	宿泊, サービス, 満足-評価語, お尋ね, 電話, 対応, 最悪

9. 形態素解析の誤り.

   形態素解析の段階で誤りが生じ，単語が正しく分割されなかったとき，不正確な素性が抽出され，誤りを引き起こすことがある．例を表 4.22 に示す．この例では、「欠品」という否定的な意見を示唆する単語の認識に失敗し，これを素性として抽出することができなかったため，評価点の推測に失敗している.

表 4.22: 形態素解析のエラーが原因となった誤りの例

項目	評価視点：部屋　　真の値：1　予測値：5
原文	主寝室にテレビがなくて別室リビングにテレビがあり、フロントに申し出ると、震災でテレビが少なくて不便をかけるとの回答。結局別室からテレビ台とも移動しテレビを見る始末。客室として欠品状態でした。
抽出したコメント	客室として欠品状態でした。
素性	客室, 欠, 状態

第5章　結論

5.1　まとめ

本研究では，レビューテキストから評価視点毎に評価点を自動的に予測する手法を提案した．まず，レビューテキストから各評価視点に関連するキーワードのマッチングにより，評価視点に関するコメントを抽出する．次に，抽出したコメントを入力として，評価点を自動的に予測するモデルを教師あり機械学習により自動構築する．機械学習アルゴリズムとしてL2正則化ロジスティック回帰を採用し，機械学習のための素性として，自立語，自立語+否定，評価語，評価語+否定の4つを用いた．楽天トラベルから，評価点が付与されたホテルのレビュー文を取得し，これを用いた評価実験を行った．ランダムに選択した500件のレビューに対し，評価視点を表わすコメントを人手でアノテーションし，それを正解データとして，コメント抽出の精度，再現率，F値を評価した．また，5分割交差検定により，提案手法による評価点推測の正解率とRMSEを調べた．コメント抽出については，そのF値は6つの評価視点について73.99%から88.91%となり，十分に高いことが確認された．また，評価点の推測については,「設備・アメニティ」以外の全ての評価視点の正解率はベースラインを上回った．また，3つの評価視点についてはRMSEも小さくなった．6つの評価視点全てについての正解率はベースラインより4.33% これらの実験結果から，提案手法の有効性が確認できた．また，評価点の推測に失敗した原因の考察も行った．その結果，一部予測不可能なデータ (評価点とレビューの内容が矛盾するなど) を除き，コメント抽出の誤り以外に，評価語辞書における評価語の不足，評価語対象の誤り，形態素解析の誤り，否定の検出の失敗などの原因が確認できた．

5.2　今後の課題

本研究の今後の課題について述べる．実験では，評価視点「設備・アメニティ」の評価点推測については，提案手法の正解率とRMSEの結果はベースラインより悪かった．原因の一つは，コメントの抽出率が低いことと考えられる．そのため，評価視点に対するコメントを抽出する手法の洗練が必要である．本研究では、評価視点に関連するキーワードを用いてコメントを抽出した．しかし,「設備・アメニティ」に関連するキーワードが数多く存在し，その内容も他の評価視点と比べて多岐にわたる．より多くのキーワードを利用するために，例えば，既存のキーワードと類似した文脈によく出現する単語を新たにキー

ワードに追加する手法を検討するべきである．多くのキーワードを用意することで，設備やアメニティに言及したコメントを漏れなく抽出できると考えられる．

また，評価点を推測する実験では，正解率が向上したのにも関わらず，評価点の誤差(RMSE)が大きくなったことがあった．4.4節の誤り分析の結果を踏まえて，既存のコメント抽出手法の洗練，否定を判別する手法の洗練，評価語辞書の拡充などを探究する．さらに，コメントの評価対象をより正確に判定する必要がある．例えば，現在の形態素解析に加え，レビュー文の文節の係り受け解析を行い，評価語と係り受け関係にある単語を評価対象とみなし，その評価対象が評価視点と一致しているかを判定する手法が考えられる．

本研究では教師あり機械学習としてL2正則化ロジスティック回帰を用いたが，学習パラメータの最適化は行わず，LIBLINEARのデフォルト値を用いていた．今後，パラメータの調節を検討する必要がある．また，現在は4種類の素性しか学習に用いていないが，評価点の推測に有効な手がかりを探究し，それを新たな素性として導入することも検討したい．

本研究では，ホテルレビューを対象に評価視点毎の評価点の予測を試みた．今後，ホテル以外の評価対象(例えばパソコンやカメラなどの製品)に対して提案手法の有効性を検証したい．また，提案手法では，評価視点はあらかじめ定義していた．しかし，ユーザによっては既存の評価視点以外の観点からの他者の評価を知りたい場合もある．例えば，本論文で想定した6つの評価視点以外の観点，例えば「部屋から見える風景」の良さを知りたいユーザもいるだろう．したがって，評価視点をあらかじめ定義するのではなく，ユーザが評価視点を入力し，それに対する評価点を推測できる方が望ましい．これも今後の重要な課題である．

謝辞

本研究を進行する際，丁寧に様々な指導を頂いた白井清昭准教授に深謝いたします．研究において，貴重な意見を頂いた池田心准教授，長谷川忍准教授に感謝いたします．また，本研究の趣旨を理解して頂き，議論などを通じて意見を頂いた白井研究室に所属する学生の皆様に感謝いたします．

参考文献

[1] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874, 2008.

[2] 東山昌彦, 乾健太郎, 松本裕治. 述語の選択選好性に着目した名詞評価極性の獲得. 言語処理学会第 14 回年次大会発表論文集, pp. 584–587, 2008.

[3] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系 CD-ROM 版. 岩波書店, 1999.

[4] Yu Jianxing, Zha Zheng-Jun, Wang Meng, and Chua Tat-Seng. Aspect ranking: Identifying important product aspects from online consumer reviews. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 1496–1505, 2011.

[5] 小林亮, 鈴木浩, 服部哲, 速水治夫. ポジティブ・ネガティブ分析を備えたユーザレビュー集約管理システムの提案. 情報処理学会研究報告, Vol. 2013-GN-87, , 2013.

[6] 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. 3, pp. 203–222, 2005.

[7] 駒田康孝, 山名早人. 商品評価ツイートからの属性語自動抽出手法の提案. In *DEIM Forum*, 第 B5-6 巻, 2014.

[8] Duc-Hong Pham, Anh-Cuong Le, and Thi-Kim-Chung Le. A least square based model for rating aspects and identifying important aspects on review text data. In *Proceedings of the 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science*, pp. 265–270, 2015.

[9] 小池惇爾, 福本文代, 松吉俊, 稲葉脩太. 評価視点別レビュー要約のための重要文候補抽出. 言語処理学会第 18 回年次大会, pp. 1188–1191, 2012.

[10] 外山洋太, 三輪誠, 佐々木裕. 文書・文間及びカテゴリ間の関係を考慮したレーティング予測. 言語処理学会第 22 回年次大会, pp. 158–161, 2016.