CE-002

宿泊予約サイトのユーザレビューにおけるクレーム抽出と内容解析による課題抽出
Complaint Sentence Detection from Customer Reviews in a Hotel Booking Site and the Content Analysis

青柳 直人† 蓮池 隆†
Naoto Aoyagi Takashi Hasuike

1. はじめに

オンラインショッピングサイトや SNS を通して，個人が商品やサービスに対する感想や意見を発信することが容易になった．このような商品やサービスに対するユーザレビューが Web 上には膨大に存在する．ユーザは商品の購入やサービスの利用を検討する際にユーザレビューを参考にすることが多く，意思決定の上で有用な情報源となっている．企業側としても，ユーザレビューから知見を得ることによって改善すべき点の発見や，マーケティングに活用することができる．したがって，ユーザレビューはユーザ側と企業側の両者にとって重要な情報である．

ユーザレビューの中でも，商品やサービスに対する不満や要求を含む「クレーム」の影響力が大きい．Laczniak ら[1]は，クレームのような否定的なレビューが購買決定にどのように影響するかに注目して実験を行い，消費者が否定的なレビューを特に重要視することを示した．したがって，企業はレビュー内のクレームに対して適切に対応する必要がある．しかし，以下の理由からレビュー内のクレームを見落としてしまう可能性がある．

(1) Web 上のレビュー数が膨大であるため，企業の担当者がすべてのレビューを読むことが困難である．
(2) 以下の図 1 のように，ユーザは一つのレビューに様々な意見を書くことが多く，その中にクレームが埋め込まれる．

> フロントの方は親切で良かったです。食事も大変美味しくいただきました。ただ、お風呂が狭いのが気になりました。全体としては満足です。

図 1 クレームが埋め込まれたレビュー例

乾ら[2]はこれらの問題に対して，レビューを文単位で扱い，ユーザの不満や要望といった内容を含むクレーム文を自動検出するモデルを提案した．このモデルでは，評価表現と文脈一貫性に基づいて教師データを自動生成し，教師データの特性を考慮してナイーブベイズモデルを拡張した．その結果，人手を介在させる場合と同等あるいはそれ以上のクレーム検出精度を示した．この研究では，クレーム文を 2 値分類問題として扱っており，クレームの内容については分析を行っていない．しかし，企業がクレームに対して適切な対応を行うには，その内容についても分析を行うことが望ましい．

宿泊施設の予約は Web 上の旅行予約サイトでの予約が主流となっている．利用者は予約サイト内の宿泊レビューを参考にして宿泊施設を選ぶことが多い．そのため，宿泊予約は他の商品やサービスと比較してレビューの重要度が高い．宿泊レビューには以下の特徴がある．

- 多くの予約サイトの宿泊レビューは実際に宿泊した人しか投稿できない．そのため，利用者になりすまして投稿することが難しく，他の業種のレビューと比較して信頼性が高い．
- 各レビューには 5 段階での評点が付いていることが多い．総合評点だけでなく，サービス，立地，部屋，食事といった項目ごとの評点も存在する．
- 宿泊施設の利用目的(レジャー，ビジネス)や，同伴者(一人，家族，恋人など)が付与されている．

辻井ら[3]は評点が同じであっても，宿泊利用者の満足度に差があり，評点にない項目でも宿泊利用者が重要視した項目が存在することを示した．このことから，宿泊施設は評点だけでなく，レビューテキストを分析することが望ましい．

本研究では，旅行に関するオンライン予約を扱う楽天トラベルのレビューデータ[4, 5]を対象に分析を行う．宿泊施設に対するレビューからクレーム文を検出するだけでなく，その内容を解析する手法を提案することで，宿泊施設にとって有用な情報を提供することを目的とする．

2. 本研究で用いる手法と既存研究

本研究で用いる代表的な手法を紹介する．2.1 節では汎用的な事前学習モデル BERT の説明を行う．2.2 節では，トピックモデルの代表的な手法を紹介する．その後，これらを用いた既存研究と本研究の立ち位置を 2.3 節で述べる．

2.1 BERT

Devlin ら[6]は自然言語処理における汎用的で高性能な事前学習済みモデルである BERT(Bidirectional Encoder Representations from Transformers)を提案した．言語理解タスク GLUE 等の自然言語処理タスクにおいて，大きく精度を向上し，最高性能を達成した．双方向の Transformer[7]をベースにしたニューラルネットワークを用いて，大量の教師なしデータで事前学習を行い，その後タスクに応じた教師ありデータでファインチューニングを行う．BERT には主に，次のように $\mathrm{BERT_{BASE}}$ と $\mathrm{BERT_{LARGE}}$ の二つの構成がある．

- $\mathrm{BERT_{BASE}}$
  L=12, H=768, A=12, パラメータ数: 1.1 億
- $\mathrm{BERT_{LARGE}}$
  L=24, H=1024, A=16, パラメータ数: 3.4 億

ここで，L は Transformer ブロックの数，H は隠れ層のサイズ，A は self-attention のヘッド数である．Transformer は翻訳タスクなどに用いられるモデルで，BERT ではこれを複数組み合わせて使う．self-attention は，入力文における離れた単語同士の依存関係を捉えるためのモデル構造である．

BERT における事前学習では次の二つのタスクを解く．

† 早稲田大学 Waseda University

① Masked Language Model (MLM)

図 2 のように，文章中の単語から 15%を無作為に選び，[MASK]トークンに置き換える．その後，前後の文脈から置き換えられる前の単語を予測する．これにより，文脈を考慮した単語の特徴ベクトルを獲得することができる．

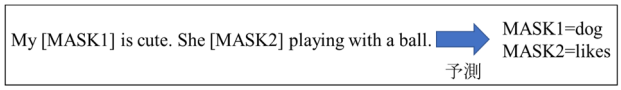

図 2　Masked Language Model の例

ただし，後述のファインチューニングでは[MASK]トークンが出現しないため，事前学習とファインチューニングの間に差異が生じてしまう．この問題を緩和するため，一部をランダムに選んだトークンで置換する，もしくはそのままにすることで対応する．

② Next Sentence Prediction (NSP)

質問応答などのタスクでは，2 文の関係性を理解することが重要である．二つの文の関係性をモデルに理解させるために，図 3 のように，2 文を入力したときに，それらが隣接文であるか否かの判定を行う．

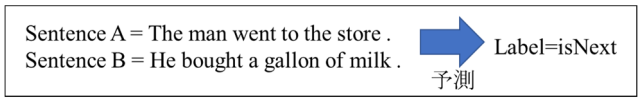

図 3　Masked Language Model の例

①および②による事前学習後のファインチューニングでは，少量の教師ありデータを使用する．BERT の出力層をタスクに合う層に付け替えることで，モデル全体を学習させる．本研究で用いる文分類では，図 4 のように入力の先頭であることを表すために使用される特殊トークン[CLS]の出力ベクトルに全結合層を追加することで，文書のラベルを予測する．

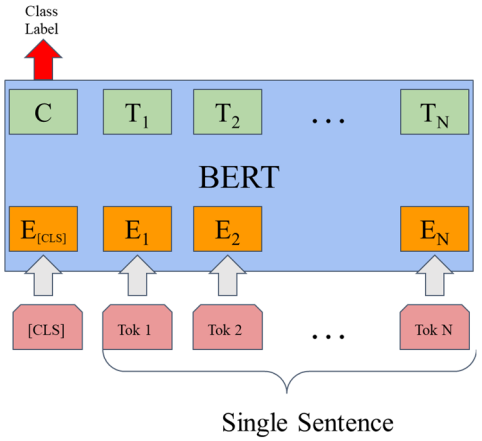

図 4　文分類におけるファインチューニング

2.2 トピックモデル

トピックモデルは大量の文書データから有益な情報を発見するための手法であり，画像処理，推薦システムなど自然言語処理以外の分野でも応用されている．トピックモデルを情報検索や文書分類などに用いることで，文書集合から潜在的なトピックを抽出し，それぞれの文書がどのようなトピックを持っているかがわかる．

トピックモデルの中の代表的なものとして，Blei ら[8]が提案した Latent Dirichlet Allocation (LDA)が挙げられる．LDA は各文書には複数の潜在トピックがあると仮定し，そのトピック分布を離散分布としてモデル化している．一方で，LDA ではトピックを抽出する際に文書単位で単語の共起性を捉えているため，ある程度の文章の長さが必要であり，Web 上に存在するユーザレビューのような短文に対しては上手く機能しないことが多い．

この問題を改善するため，Yan ら[9]は短文のためのトピックモデルである Biterm Topic Model (BTM) を提案した．BTM では，文書全体における単語の共起パターンの生成過程をモデル化することで，短文で発生するスパース性の問題を解決する．

BTM の文書集合の生成過程は以下のようになる．ここで，B はバイタームの集合を表し，$|B|$はその総数を示す．

STEP1：K 個のトピックに対して，$\boldsymbol{\phi}_k \sim Dirichlet(\boldsymbol{\beta})$を選択

STEP2：コーパスに対して，$\boldsymbol{\theta} \sim Dirichlet(\boldsymbol{\alpha})$を選択

STEP3：バイターム$b \in B$に対して，

(3-1) トピック$z \sim Multinomial(\boldsymbol{\theta})$を選択

(3.2) 単語$w_i, w_j \sim Multinomial(\boldsymbol{\phi}_z)$を選択

単語分布$\boldsymbol{\phi}_k$とトピック分布$\boldsymbol{\theta}$は確率ベクトルであるので Dirichlet 分布，トピック z と単語w_i, w_jは離散値なので多項分布を仮定している．BTM のグラフィカルモデルを図 5 に示す．

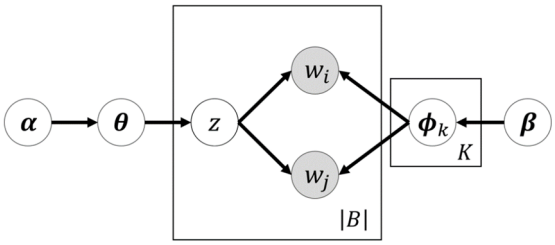

図 5　BTM のグラフィカルモデル

2.3 既存研究

本研究に関連するいくつかの既存研究について，その内容と課題や改善点を述べる．

2.3.1 教師データ自動生成によるクレーム検出

乾ら[2]は宿泊レビューからクレームが記述された文(クレーム文)を自動検出するタスクに対して，評価表現と文脈一貫性に基づく教師データ自動生成と，生成されたデータの性質を踏まえてナイーブベイズモデルを拡張したモデルを提案した．

この研究では、各文がクレームを表している【クレーム】か，クレームを表していない【非クレーム】のどちらかのラベルを付ける．まず，あらかじめ定義した評価表現辞書に基づいてラベル付けを行う．評価表現辞書とは，「良い」「悪い」等の，評価対象に対する評価の明示的な表現を集めた辞書である．この段階で付けたラベルを核文ラベルと呼ぶ．また，評価表現が連続する文脈が形成される傾向を

利用して近接文ラベル付けを行う．これは，核文ラベル付けで考慮した評価表現を含む文の周辺文について，「評価表現の存在に基づいてクレーム文として選ばれた文の前後文脈に位置する文は，やはりクレーム文である」と仮定し，核文の周辺文についてラベル付けを行う．

次に，自動作成したデータの性質を考慮してナイーブベイズモデルを拡張する．通常のナイーブベイズモデルでは，文 s のクラス c を予測する際，以下の式で決定する．

$$\begin{aligned}\hat{c} &= argmax_c P(c|s)\\ &= argmax_c P(c)P(s|c)\\ &= argmax_c \left\{ \log p_c + \sum_{w \in V} n_w(s) \log q_{w,c} \right\}\end{aligned}$$

ここで，V は単語の集合，$n_w(s)$は文 s における単語 w の出現回数である．$q_{w,c}$とp_cは以下の式で計算される．

$$q_{w,c} = \frac{n_{w,c}(D) + 1}{\sum_w n_{w,c}(D) + |V|}$$

$$p_c = \frac{n_c(D) + 1}{\sum_c n_c(D) + |C|}$$

D は教師データの文集合，$n_{w,c}(D)$は D においてクラス c に属する文の単語 w の出現回数，$n_c(D)$)は D においてクラス c に属する文の数である．

この通常のナイーブベイズモデルに対して文脈を考慮した拡張をする．拡張モデルでは以下の式でクラスを求める．

$$\begin{aligned}\hat{c} &= argmax_c P(c)P(s|c)\\ &= argmax_c \left\{ \log p_c + \sum_w n_w(s) \log q_{w,c}^{tgt} \right.\\ &\left. + \frac{1}{|ctx(s,N)|} \sum_w n_w\big(ctx(s,N)\big) \log q_{w,c}^{ctx} \right\}\end{aligned}$$

$ctx(s,N)$は文 s の周辺文脈に位置する前方および後方のそれぞれ N 文から構成される文の集合，$n_w\big(ctx(s,N)\big)$は集合$ctx(s,N)$に属する文における単語 w の出現回数である．ここで，p_c，$q_{w,c}^{tgt}$，$q_{w,c}^{ctx}$は次の式で求める．

$$p_c = \frac{n_c(D_{tgt}) + 1}{\sum_c n_c(D_{tgt}) + |C|}$$

$$q_{w,c}^{tgt} = \frac{n_{w,c}(D_{tgt}) + 1}{\sum_w n_{w,c}(D_{tgt}) + |V_{tgt}|}$$

$$q_{w,c}^{ctx} = \frac{n_{w,c}(D_{ctx}) + 1}{\sum_w n_{w,c}(D_{ctx}) + |V_{ctx}|}$$

D_{tgt}は核文ラベル付きデータ，D_{ctx}は近接文ラベル付きデータである．このモデルを NB+ctx と呼ぶ．文脈が考慮でき，クレーム文検出精度の向上することが可能となった．

一方で，残された課題として，クレーム文検出精度のさらなる向上，およびクレーム文か非クレーム文の 2 値分類だけでなく，クレーム内容についても自動分類することが挙げられている．

2.3.2 トピックモデルによるクレーム解析

Hu ら[10]は，トピック間に相関を想定した上で文書の不随情報を考慮することができるトピックモデル Structural Topic Model (STM)[11]によって宿泊施設利用者のクレームを解析した．ニューヨークのホテルに対するレビューを対象に分析を行い，利用者のクレーム原因への理解を試みた．その結果，5 段階の評点が 1~2 点のネガティブなレビューに多く存在する 10 トピックを明らかにした．さらに，利用者のクレームがホテルのグレードごとにどのように異なるのかを調べた．グレードの高いホテルに対してはサービスに関する問題，グレードが低いホテルに対しては設備に関する問題が多く書かれていることがわかった．

この研究は，レビュー内のクレーム内容を詳細に分析している．一方で，レビューを文単位で扱っておらず，1 節で述べたような高い評点をつけていたとしてもクレームが埋め込まれているようなレビューを抽出することができない．そのため，クレームの見落としを防ぐためにレビューを文単位で分析するという課題が残されている．

2.3.3 Aspect-Based Sentiment Analysis

レビューの解析では，内容が肯定的か否定的かを分類する Sentiment Analysis に関する研究が盛んに行われてきた．Sentiment Analysis には，何が肯定的・否定的に評価されているのかがわからないという問題がある．そこで近年，何が肯定的もしくは否定的に評価されているのかまで知ろうとする Aspect-Based Sentiment Analysis[12]と呼ばれるタスクについての研究が増加している．図 6 に具体例を示す．

朝食は美味しかったです。
Food: Positive
フロントの方が不親切だったのが残念でした。
Service: Negative

図 6　Aspect-Based Sentiment Analysis の例

1 文目では，朝食について言及されているので aspect は Food，評価は肯定的なものなので Positive と判定されている．2 文目では，フロントの接客について言及されているので aspect は Service，評価は否定的なものなので Negative と判定されている．このように Aspect-Based Sentiment Analysis では，予め定義した aspect に該当する部分を特定し，それに対して Sentiment Analysis を行う．

しかし，現状では，特にデータが少ない aspect に対して正確に分類を行うことは困難である．また，解析者があらかじめ aspect を決めておく必要があり，データから新たな知見が得られることは少ない．そのため，Aspect-Based Sentiment Analysis によってクレーム文を網羅的に抽出し，その内容を解析することは難しい．

3. 提案手法

本研究では，前節と同様に宿泊施設に対するレビューからクレーム文を検出することに加えて，検出した文の内容を分析するモデルを構築する．2.3.3 節で説明した Aspect-Based Sentiment Analysis によって文の極性と aspect の分類を同時にできれば，極性の低い文の aspect を調べることでクレーム内容を把握することができる．しかし，現状では aspect 分類の精度が十分でなく，クレーム内容を正確に分類することができないため，Aspect-Based Sentiment Analysis による分析は困難である．また，2.3.2 節のようにトピックモデルでクレーム内容を特定することは有用であ

るが，レビューを一つの文書として扱ってしまうと，レビュー内に埋め込まれたクレームを見落としてしまう問題を回避できない．

そこで，本研究では以下の 2 段階のアプローチによって宿泊レビューの分析を行う．まず，2.3.1 節のようにレビューを文単位で扱い，クレーム文を検出する．その後，検出したクレーム文について，トピックモデルを適用することでクレーム内容の解析を行う． 1 段階目のクレーム文検出には BERT，2 段階目のクレーム内容解析には BTM をもとにした手法を提案する．

3.1 クレーム文検出

まず，レビューを文ごとに分割し，各文がクレーム文か否かを判定する．ここで，本研究ではクレーム文を「商品やサービスに対しての不満を表す苦情，もしくは不満を解消するための要望の内容が含まれる文」と定義する．この定義から，1 文に商品やサービスに対して肯定的な内容とクレーム内容の両方が含まれている場合でも，クレーム文となる．これは，宿泊施設としてはクレーム内容と同時に褒めの内容が書かれていたとしても，利用者が不満に感じた点は把握しておく必要があると考えられるためである．

本研究では BERT によってクレーム文を検出することを考える．BERT を選択した理由として．一つには，多くの文書分類タスクにおいて高い精度を出していること．もう一つには，各文がクレームか否かのラベルが付いているデータが入手困難なため，少量のラベル付きデータでも機能するモデルが好ましいことが挙げられる．BERT は大量の教師なしデータで事前学習を行っているため，ラベル付きデータが少量であっても高い分類性能を示す．

通常，BERT における文書分類のファインチューニングでは，2.1 節で示したように，[CLS]トークンに対応する出力ベクトルを全結合層にかける．しかし，[CLS]トークンに対応する出力ベクトルは文全体の意味を捉えた特徴ベクトルとなっているため，通常のファインチューニングを行った場合，文全体としてクレーム文か否かを判定してしまう可能性が高い．したがって，文中に肯定的な表現が多いがクレーム内容も含んでいる文を見落としてしまうことが考えられる．このような問題を解消するために，図 6 のようなファインチューニングを考える．

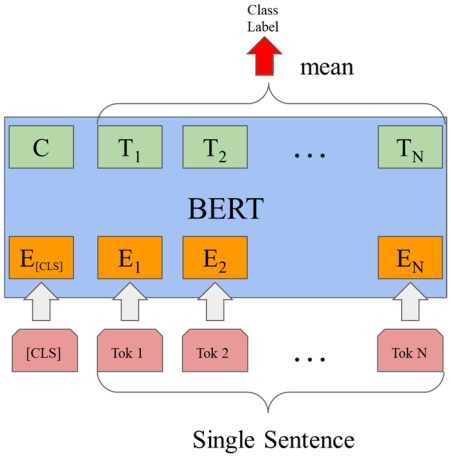

図 6　各トークンの平均ベクトルを用いたファインチューニング

ファインチューニングに[CLS]トークンではなく，各トークンに対応する出力ベクトルの平均を用いることで，前述のような問題の解決を試みる．これにより，文内でクレーム内容を述べた部分の単語に対応する出力ベクトルの特徴を残すことができ,肯定的な表現が多い場合でもクレーム内容が含まれていればクレーム文と判定することができると考えられる． 以降，図 7 のファインチューニングを BERT+mean と呼ぶ．

3.2 内容解析

次に，抽出したクレーム文の内容解析を行う．本研究では，レビューを文単位で扱っているため，抽出したクレーム文は短文である．クレーム文からトピックを抽出するために，短文においても機能するトピックモデルである BTM を用いる．

BTM によってクレーム文のトピックを抽出した後，各宿泊施設が競合する宿泊施設と比較してどのようなクレームが多いかを明らかにする．これにより，宿泊施設が優先的に解決すべき問題を抽出することができる．

以下，競合との比較方法について説明する．BTM から文 s でのトピック z の構成比率は次のようになる．

$$P(z|s) = \sum_b P(z|b)P(b|s)$$

$P(z|s)$はベイズの定理から，以下のように計算できる．

$$P(z|b) = \frac{P(z)P(w_i|z)P(w_j|z)}{\sum_z P(z)P(w_i|z)P(w_j|z)}$$

ここで，$P(z) = \theta_z, P(w_i|z) = \phi_{i|z}$である．残りの$P(b|s)$は以下の式で計算される．

$$P(b|s) = \frac{n_s(b)}{\sum_b n_s(b)}$$

$n_s(b)$は文 s におけるバイターム b の出現回数である．$P(z|s)$を用いて宿泊施設 h のトピック分布を計算する．ここで，s_hは宿泊施設 h に対するクレーム文，n_hはその数である．

$$P(z|h) = \frac{\sum_{s_h} P(z|s_h)}{n_h}$$

また，宿泊施設 h におけるトピック z のリスクを以下のように定義する．

$$Risk_z^h = \frac{P(z|h)}{\sum_{h \in H} P(z|h)/|H|}$$

ここで，H は宿泊施設の集合で，$|H|$はその数である．$Risk_z^h$は宿泊施設 h におけるトピック z のクレームが，分析対象の宿泊施設の平均と比較してどれだけ多いかを表す．よって，$Risk_z^h > 1$ならば，トピック z のクレームが多いと判断することができる．リスクが大きいほど，宿泊施設はそのトピックに関する問題を早急に解決する必要がある．

4. 実データ分析

実データを対象に分析を行い，提案手法の有効性を検証し，分析の結果とそれに対する考察を示す．

4.1 データセット

楽天トラベルデータ[4]から 2017 年から 2019 年までの 3 年分のデータ 643 万 2,538 件を使用した．また，クレーム文出精度の確認のために，筑波大学文単位評価極性タグ付きコーパス (TSUKUBA コーパス)[5]を用いた．このデータは，楽天トラベルのレビューデータに対して，文単位で評価極性情報を付与したコーパスである．データ数は，レビュー1,000 件（4,309 文）である．評価情報には，褒め・苦情・要求・ニュートラル・評価なし・その他/保留の 6 種類のラベルがあるが，苦情または要求が付与されている文をクレーム文とした．

4.2 分析条件

形態素解析は形態素解析エンジン MeCab[19]を使用した．MeCab のシステム辞書には，固有名詞の解析に強い mecab-ipadic-NEologd[20]を用いた．品詞は名詞，動詞，形容詞に限定した．前処理として単語の正規化，ストップワードの除去をした．また，レビューの文分割には日本語 NLP ライブラリ GiNZA[21]を用いた．

4.3 クレーム文検出の結果と考察

BERT を用いたクレーム文検出を行うために，ファインチューニング用の教師データを作成する．まず，2018 年度のレビューデータから，無作為に 1,000 件のレビューを選ぶ．レビューを文単位に分割し 4,242 文を作成した．そして，作業者に 1 文ごとに分割されたレビューを提示し，各文が褒め，苦情，要求，ニュートラル，その他のラベルを付与するアノテーションを行った．複数選択可能となっており，1 文に複数のラベルが割り当てられることがある．このデータから，苦情または要求のラベルが付与されている文をクレーム文とした．

用意した教師データで，BERT のファインチューニングを行う．BERT モデルとして，学習済み日本語 BERT モデル[22]を用いた．これは，tokenizer に MeCab- NEologd，コーパスは日本語版 Wikipedia を用いて学習を行っている BERT モデルである．TSUKUBA コーパスをテストデータとして，クレーム検出の精度を確認する．その際，3.1 節で説明した各トークンに対応する出力ベクトルの平均を用いる方法 (BERT+mean)に加えて，各トークンに対応する出力ベクトルに対して各要素の最大値を取ったベクトルを用いる方法 (BERT+max)，BERT+mean と BERT+max のベクトルを concat したベクトルを用いる方法 (BERT+concat)の検証も行う．また，評価極性辞書による手法 (dictionary-base)，通常のナイーブベイズモデル (NB)，3.1 節で説明した先行研究の手法 (NB+ctx)と比較する．結果を表 1 に示す．

表 1　クレーム文検出精度

Model	AUC	Accuracy	Precision	Recall	F1
dictionary-base	-	0.777	0.644	0.532	0.582
NB	0.821	0.817	0.733	0.587	0.652
NB+ctx	0.887	0.731	0.523	**0.914**	0.665
BERT	0.931	0.861	0.781	0.732	0.755
BERT+mean	**0.936**	**0.883**	**0.819**	0.772	**0.795**
BERT+max	0.918	0.856	0.755	0.750	0.752
BERT+concat	0.919	0.864	0.808	0.702	0.751

他の手法と比較して，BERT+mean を使うことで AUC, Accuracy, Precision, F1 が改善していることがわかる．

次に，閾値を変化させたときの精度を確認するために図 7 に ROC 曲線，図 8 に precision-recall 曲線を示す．

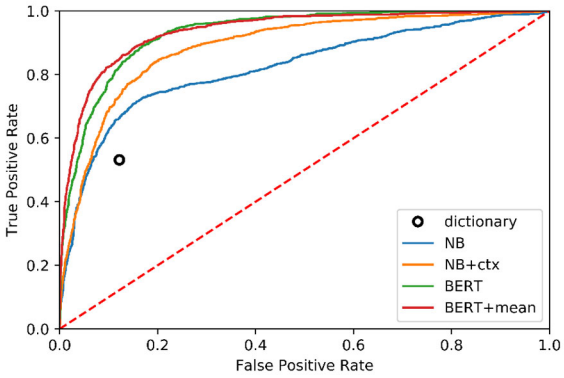

図 7　ROC 曲線

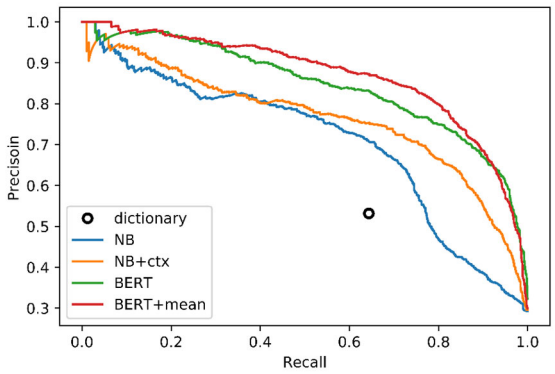

図 8　precision-recall 曲線

ROC 曲線は，文がクレーム文であると予測する閾値を 1 から 0 に動かし，そのときの偽陽性率を横軸，真陽性率を縦軸にとって描くことができる．このようにして描いた ROC 曲線の下部の面積が AUC である．曲線がグラフ左上の近い位置にあるほどモデルが優れていることを表す．図 7 から，閾値を変化させたときも BERT+mean の曲線が左上にあり，高い分類精度を確認できた．また図 8 の precision-recall 曲線は，横軸に Recall，縦軸に Precision をとった曲線であり，曲線が右上に近い位置にあるほどモデルが優れていることを表す．この図からも BERT+mean が最も良い分類精度を示していることがわかる．

BERT+mean を使うことで，どのような文を正確に分類できたかを調べる．そこで，通常の BERT のファインチューニングでは誤分類だが，BERT+mean で正解だった例を表 2 に示す．表 2 から，クレーム文では「良かった」という褒めの内容も含まれていることがわかる．通常のファインチューニングでは，文全体での意味を捉えていることからこのような文を非クレーム文と誤分類してしまっていると考えられる．BERT+mean を用いることで，「ホスピタリティの問題」や「わかりずらかった」などのクレームに該当する部分の特徴を残すことができているため，このようなクレーム文でも正確に分類できていると考えられる．

表 2　BERT+mean によって正確に分類できた例

ラベル	レビュー文
クレーム	お部屋からの眺めは海外のようでとても良かったですし、ホテルの設備も良かったのですが、なんと言ってもホスピタリティの問題が。。。
クレーム	駐車場の場所がわかりずらかったですが、駐車場の従業員の方の対応が良かったです。
非クレーム	実際フロントの対応、客室係りの対応、申し分ない対応でした。
非クレーム	夜は静かでしたが、トータルでは可もなく不可もなくといったところです。

また，非クレーム文に関しては，「申し分ない」「可もなく不可もなく」といった言葉が含まれている文が通常の BERT で誤分類されている．一方で，BERT+mean では，文中の単語に対応するベクトルの平均をとるため，否定的な特徴が緩和され，誤分類を減らすことができたのではないかと考えられる．

また，ファインチューニングで用いるデータ量を変化させたときのクレーム文検出精度への影響を調べる．図 9 は学習データ数を 100 件ずつ増やしたときの AUC である．実線が BERT+mean における AUC の推移、点線が 0.9 である．

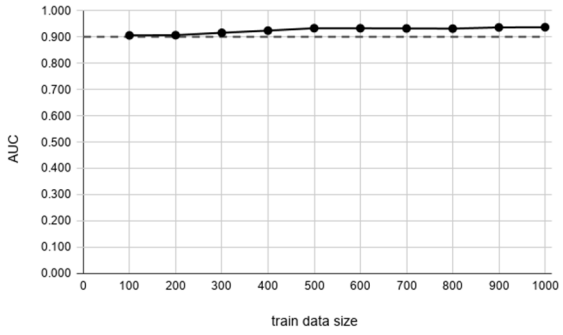

図 9　学習データ量の影響

学習データ数が 100 件と少量の場合でも 0.9 を超えている．これは，事前学習により良い文脈付き特徴ベクトルを獲得できているためであると考えられる．

次に，BERT を用いた手法の精度を詳細に調べるために，TSUKUBA コーパスを使用して 10 分割交差検証を行う．その結果を表 3 に示す．

表 3　10 分割交差検証によるクレーム文検出精度

Model	AUC	Accuracy	Precision	Recall	F1
BERT	0.937	0.871	0.780	0.780	0.780
BERT+mean	**0.948**	**0.884**	**0.794**	0.825	**0.807**
BERT+max	0.927	0.863	0.750	0.808	0.776
BERT+concat	0.938	0.870	0.755	**0.837**	0.791

10 分割交差検証の結果でも，BERT+mean が最も良い検出精度となった．

このモデルを用いて，2017 年から 2019 年のデータからクレーム文を抽出する．その際，Recall が 0.9 以上となるように分類時の閾値を調整する．これは，宿泊施設にとって，非クレーム文を誤検出が増加したとしても，クレーム文の見落としを減らすことが重要だと考えられるからである．閾値調整後の分類精度は，Accuracy=0.850, Precision=0.685, Recall=0.901, F1=0.778 となった．このモデルを使ってクレーム文検出を行った．その結果，2,332,828 文 / 6,432,538 文 (36.3%) がクレーム文として検出された．

4.4 内容解析の結果と考察

前節でクレーム文と判定された文を対象に内容解析を行う．地域や宿泊施設の種類ごとに利用者が宿泊施設に対して求めることが異なると考えられる．そこで，宿泊施設のエリアと，利用者の利用目的(ビジネス，レジャー，その他)によって対象を絞る．以下，①宿泊施設のエリアが東京都，利用目的がビジネスであるレビューと，②宿泊施設のエリアが京都府，利用目的がレジャーであるレビューを対象とした分析例を示す．BTM のパラメータは，α=50/K, β=0.01, K=30，バイターム取得時のウィンドウサイズを 15 として，ギブスサンプリングで推論を行う．

4.4.1 東京・ビジネス利用の場合

東京のホテルを対象としたレビューで，ビジネス目的であるものを分析する．対象としたデータは，レビュー件数：38,797，文数：89,690，宿泊施設数：1,221 である．

BTM で得られた 30 トピックのうち，代表的なものを表 4 に示す．それぞれのトピックにおいて出現確率が高い 10 単語を上から順に表示している．

表 4　代表的なトピック (東京・ビジネス利用)

topic4 対応	topic8 空調	topic18 価格	topic23 立地	topic28 におい	topic29 食事
フロント	エアコン	宿泊	駅	部屋	朝食
時	部屋	料金	ホテル	臭い	残念
チェックイン	寒い	ホテル	近く	禁煙	食べる
部屋	温度	利用	近い	喫煙	良い
予約	空調	値段	歩く	気	食事
ホテル	暑い	高い	コンビニ	ルーム	少ない
言う	暖房	安い	便利	タバコ	種類
対応	設定	予約	立地	匂い	無料
宿泊	調整	円	良い	室	バイキング
電話	日	価格	分	予約	時

上位の単語から判断してトピックに解釈を与えた．例えば，トピック 4 については，”フロント”，”チェックイン”，”対応”などの単語が上位にあるため，スタッフの「対応」に関するトピックであると解釈した．表 4 に掲載していないトピックにも同様に解釈を与えた結果，30 トピックのうち，28 トピックは解釈を与えることができた．

次に，特定の宿泊施設の立場から，他の競合と比較してどのようなクレームが多いのかを明らかにするため，ホテル A を対象として分析を行う．ホテル A についてリスクを計算した結果を図 10 に示す．ここで，比較対象の宿泊施設集合 H は同一エリアの同一価格帯の 64 ホテルとした．図 10 からトピック 6 (混雑)，トピック 7 (エレベーター)，トピック 15 (時間帯)，トピック 30 (広さ)のリスクが 1.5 以上と非常に高く出ている．具体的にこれらのトピックにはどのようなクレームがあるか調べる．

例えば，表 5 にトピック 6 (混雑)に属するクレーム文を一部示す．クレーム内容から，外国人利用者のチェックイン・チェックアウトが詰まってしまうことで混雑が生じていることがわかる．ホテル A はビジネス利用者だけでなく，観光・レジャーなどで利用する客も多くいるため，このような問題が生じてしまっている．必要事項を入力すれば自

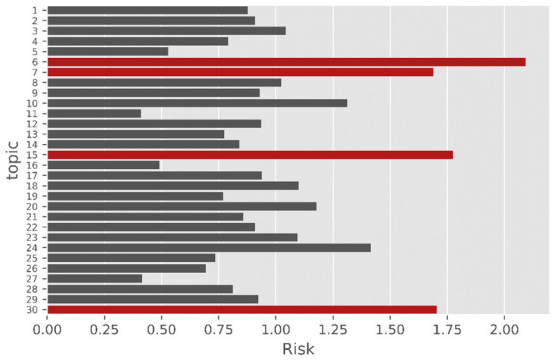

図 10　ホテル A のリスク

表 5　トピック 6 (混雑)に属するクレーム文

topic	クレーム文
6	スマートチェックインが複雑なためスタッフ介助が必要になっている。
6	海外の方が多いためか、チェックインにもチェックアウトにも時間がかかり、困ったことが一番のマイナス点ですね。
6	エレベーターはインバウンドで占拠されてなかなか乗れなかったが、概ね満足しました。

動でチェックインすることができる機械を設置するなどの対応が考えられる．

次に，表 6 にトピック 30 (広さ)に属するクレーン文を一部示す．

表 6　トピック 30 (広さ)に属するクレーム文

topic	クレーム文
30	ただ、部屋が狭いのだけが難点。
30	お任せでしたので、角部屋で狭いのは仕方ないところでした、歯ブラシは欲しかったです。
30	１点、気になる点としては机がオシャレなのですがマウスをおけるスペースが無い小さいものだったのでちょっと仕事がしにくいサイズでした。
30	通路が狭くて歩きにくい。

クレーム内容を見ると，主に部屋が狭いことが問題になっていることがわかる．また，机が小さく，仕事をしづらいというさらに，通路が狭いということも利用者の不満を感じる原因となっている．部屋の広さや通路の幅を変えることは容易ではないため，レイアウトの見直しを行うことで少しでも広く感じられるようにすることが望ましい．また，ビジネス利用を考えると，部屋で仕事をすることも十分考えられるので，マウスと PC を置ける程度の大きさの机に取り換えることを検討すべきである．

4.4.2 京都・レジャー利用の場合

京都府のホテルを対象としたレビューで，旅行の目的がレジャー，その他であるものを分析する．対象としたデータは，レビュー件数：20,354，文数：54,559，宿泊施設数：1,104 である．

BTM で得られた 30 トピックのうち，代表的なものを表 7 に示す．それぞれのトピックにおいて出現確率が高い 10 単語を上から順に表示している．

表 7　代表的なトピック (京都・レジャー利用)

topic6 騒音	topic8 食事	topic18 接客	topic24 眺め	topic28 解釈困難	topic30 予約
音	朝食	方	部屋	くる	予約
気	食べる	スタッフ	窓	時	時
部屋	美味しい	対応	見える	言う	フロント
聞こえる	食事	フロント	残念	食事	部屋
声	良い	良い	外	行く	宿泊
壁	バイキング	ホテル	良い	部屋	チェックイン
人	パン	時	お部屋	朝食	言う
廊下	残念	感じ	階	料理	電話
響く	種類	男性	景色	食べる	確認
時	多い	女性	開ける	来る	ホテル

上位の単語から判断して，トピックに解釈を与えた．例えば，トピック 6 については，”音”，”部屋”，”声”などの単語から判断して，部屋での「騒音」に関するトピックであると解釈した．一方で，トピック 28 のように解釈が困難であるトピックも見られたが，30 トピック中，27 トピックが解釈可能であった．

次に，特定の宿泊施設の立場から，他の競合と比較してどのようなクレームが多いのかを，京都東山にあるホテル B を対象にして明らかにする．ホテル B についてリスクを計算した結果を図 11 に示す．ここで，比較対象の宿泊施設集合 H は，エリアと価格帯が同一の 68 施設とした．

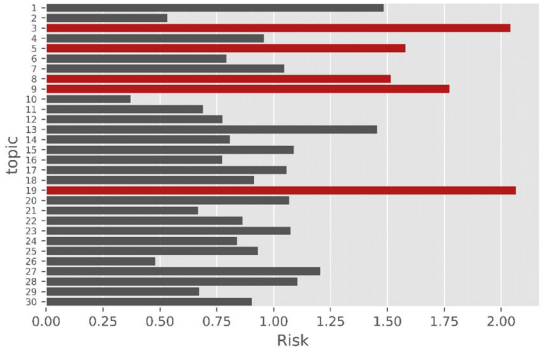

図 11　ホテル B のリスク

図 11 からトピック 3 (空調)，トピック 5 (風呂)，トピック 8 (食事)，トピック 9 (におい)，トピック 19 (掃除)のリスクが高いことがわかる．この中でも，リスクが 2 以上と非常に高いトピック 3 (空調)とトピック 19 (掃除)について，具体的なクレーム内容を調べる．

表 8 にトピック 3 (空調)に属するクレーム文を一部示す．

表 8　トトピック 3 (空調)に属するクレーム文

topic	クレーム文
3	部屋の乾燥がひどく加湿器があればと思いました。
3	ただ 空調の設定が出来ず少々暑かったです。
3	しかし室内調整が殆んどできず「弱」でも寒かった。
3	部屋の冷房がオンオフのみなので、寒くなったら止める、暑くなったら入れるの繰り返しが面倒です。

空調の調整ができないことにより，「暑い」「寒い」といった不満が生じていることがわかる．空調のシステムが故障しているか，利用者が調整できないようなシステム上の

制限をかけてしまっている恐れがある．ホテル B はこのような問題を解決し，利用者自身で空調の設定温度や風量を調整できるようにすることが求められる．

次に，表 9 にトピック 19（掃除）に属するクレーム文を一部示す．

表 9　トトピック 19 (掃除)に属するクレーム文

topic	クレーム文
19	部屋に入ってすぐに床に透明なゴミが落ちてました。
19	髪の毛が散見された。
19	洗面所の照明のカバーの上には埃がびっしりで、何ヵ月も掃除してないのでは？
19	布団に長い髪の毛も。

抽出されたクレーム文から，布団に髪の毛が落ちていたり，部屋にゴミが落ちていたりといった問題が利用者の不満になっていることがわかる．掃除されているかをチェックする仕組みを導入するなどの対応が考えられる．

4.5 結果の考察

以上の結果から，それぞれの宿泊施設のリスクを計算することによって，問題点を抽出できることが確認できた．宿泊施設はリスクが高いトピックに属する文をチェックすることで，利用者のクレームを把握することができる．しかし，実際にはクレーム文でない文も含まれているため，リスクを計算する際にノイズとなっている可能性がある．そのため，クレーム文出精度を高めることで，より正確に宿泊施設の問題点を抽出することが可能であると考えられる．また，リスクが高いからといって必ず重要な問題点が含まれているわけではないことに注意する必要がある．トピックが同一であってもその内容がそれぞれ異なる場合，一つ一つのクレームの重要度が高い訳ではない．よって，宿泊施設の担当者はリスクの高いトピックから順にクレーム文を確認し，内容を把握することが望ましい．

5. まとめと今後の課題

本研究では，宿泊予約サイトのレビューからクレーム文を検出し，その内容を解析する手法の構築を試み，宿泊施設にとって，競合施設と比較したときに，どのようなトピックに関するクレームが多いのかを定量的に測る指標を提案した．また，宿泊予約サイトのレビューデータに提案手法を適用し，提案手法の有効性を評価した．クレーム文検出においては，従来手法よりも高い精度で分類できることを示した．教師ありデータが少量であっても，事前学習済みモデル BERT を用いることでクレーム文を検出することができた．さらに，内容解析では東京・ビジネス利用，京都・レジャー利用のレビューに含まれるクレーム文を分析した．特定の宿泊施設の立場から解析を行うことで，利用者がどのような点に不満や要望を抱いているのかを明らかにした．

今後の課題として，まず宿泊レビューでの事前学習が考えらえる．本研究では，日本語版 Wikipedia を用いて学習を行った BERT モデルでクレーム文検出を行った．しかし，宿泊レビュー特有の言い回しや単語があった場合，事前学習でそれらの知識をモデルに与えることができない．したがって，Wikipedia での事前学習に加えて，宿泊レビューデータを用いて事前学習を追加で行うことでクレーム文検出精度を向上することができると考えられる．さらには，件数は少ないが重要なレビューの抽出方法の検討も挙げられる．内容解析において，競合する宿泊施設と比較してクレーム文の多い内容が重要であるとした．しかし，本来は同じ内容のクレームが増加する前に対応することが望ましい．そのため，これまで言及されてこなかった内容のクレーム文を検知できるように，時系列での内容解析が必要である．

謝辞

本研究では，国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」を利用しました．貴重なデータのご提供に深く感謝いたします．

参考文献

[1] R.N. Laczniak, T.E. DeCarlo, and S.N. Ramaswami, “Consumers’ responses to negative word-of-mouth communication: An attribution theory perspective”, Journal of consumer Psychology, Vol. 11, No. 1, pp. 57–73 (2001).
[2] 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, Vol. 20, No. 5, pp. 683‐705 (2013).
[3] 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方法”, デジタルプラクティス, Vol. 3, No. 4, pp. 289‐296 (2012).
[4] 楽天グループ株式会社, 楽天トラベルデータ, 国立情報学研究所情報学研究データリポジトリ. (データセット), https://doi.org/10.32130/idr.2.2, 2020
[5] 楽天グループ株式会社, 筑波大学文単位評価極性タグ付きコーパス, 国立情報学研究所情報学研究データリポジトリ. (データセット), https://doi.org/10.32130/idr.2.6, 2014
[6] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv:1810.04805 (2018).
[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need”, Advances in neural information processing systems, Vol. 30, pp. 5998–6008 (2017).
[8] D.M. Blei, A.Y. Ng, and M.I. Jordan, “Latent dirichlet allocation”, Journal of machine Learning research, Vol. 3 No. Jan, pp. 993‐1022 (2003).
[9] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A biterm topic model for short texts”, In Proceedings of the 22nd international conference on World Wide Web, pp. 1445–1456 (2013).
[10] N. Hu, T. Zhang, B. Gao, and I. Bose, “What do hotel customers complain about? text analysis using structural topic model”, Tourism Management, Vol. 72, pp. 417–426 (2019).
[11] M.E. Roberts, B.M. Stewart, D. Tingley, and E.M. Airoldi, “The structural topic model and applied social science”, In Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, Vol. 4 (2013).
[12] B. Liu, “Sentiment analysis and opinion mining”, Morgan & Claypool Publishers (2012).
[13] T. Kudo, K. Yamamoto, and Y. Matsumoto, “Applying conditional random fields to Japanese morphological analysis”, In Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 230–237 (2004).
[14] T. Hashimoto, T. Sato, and M. Okumura, “Implementation of a word segmentation dictionary called mecab-ipadic-neologd and study on how to use it effectively for information retrieval (in Japanese)”, In Proceedings of the Twenty-three Annual Meeting of the Association for Natural Language Processing, NLP2017–B6–1 (2017).
[15] 松田寛, 大村舞, 浅原正幸, “短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習”, 言語処理学会第 25 回年次大会発表論文集, pp. 201‐204 (2019).
[16] Pretrained Japanese BERT models, https://github.com/cl-tohoku/bert-japanese.