

Title	複数の属性に対する評価を含む宿泊施設レビューに対する多様な返信の自動生成
Author(s)	村越, 裕太
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18895
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 修士(情報科学)

概要

インターネットの普及に伴い、宿泊施設の予約がウェブサイトを通じて行われるようになった。また、多くのオンライン予約サイトで、ユーザによるレビューの投稿機能とそれに対する宿泊施設の返信機能が提供されている。ユーザのレビューに返信することは宿泊施設にとってブランドイメージを保護し、顧客満足度を高める効果がある。一方で、大量のレビューに対応することは多くの時間とリソースを要する作業であり、特に否定的なレビューへの返信はスタッフにかかる精神的な負担が大きい。このため、宿泊施設レビューへの返信を自動生成するシステムが求められている。

レビューに対し返信を自動生成する既存の研究の多くは、アプリケーションレビューや商品レビューを対象にしており、宿泊施設のレビューに対する返信に焦点を当てた研究はあまり多くない。加えて、これらの研究ではレビュー内の複数の評価対象の属性に対して網羅的に言及する返信の生成には留意していない。ここで宿泊施設の「属性」とは、「部屋」「風呂」「スタッフ」など、評価の対象となる宿泊施設の設備・機能・サービスなどを指す。ユーザが複数の属性に対して不満を表明しているにも関わらず、その一部についてだけ返信をすると、ユーザは不誠実な対応と感じ、宿泊施設に対して悪い印象を持つ可能性がある。

本研究は、宿泊施設に関する低評価レビューに対し、適切な返信を生成することを目的とする。この際、以下の2つの点に留意する。ひとつは、ユーザが宿泊施設に関する複数の属性に対して不満を表明しているとき、その全ての属性に言及することである。もうひとつは多様な返信の生成である。当たり障りのない一般的な表現を生成するだけでなく、ユーザの不満に応じた適切な表現を生成することを目指す。

提案手法の手順は次の通りである。初めに、レビューを個々の文に分割する。次に、苦情判定モデルにより、各レビュー文が苦情であるかを判定し、苦情でない文は排除する。続いて、苦情と判定されたレビュー文を返信生成モデルへ入力し、返信を生成する。最後に、各レビュー文に対して生成された返信を統合し、最終的な返信を出力する。

苦情判定モデルは、Bidirectional Encoder Representations from Transformers (BERT) により学習する。楽天トラベルのデータセットから、苦情ラベルが付与され、かつ単一の文から構成されるレビューを正例として抽出する。また、苦情ラベルが付与されていない単文のレビューを正例と同じ数だけランダムにサンプリングし、負例とする。これを訓練データとしてBERTをファインチューニングする。

返信生成モデルは、系列変換モデルである Bidirectional Auto-Regressive Transformer (BART) をファインチューニングすることで学習する。訓練データとして、楽天トラベルのデータセットから取得したレビューとそれに対する宿泊施設の返信の組を用いる。また、訓練データの品質を高めるために、2つの訓練データのフィルタリング手法を採用する。1つ目は、属性に言及しない返信のフィルタリングである。レビューと返信の両方に属性語が存在する組のみを残し、それ以外は

削除する。これにより、返信生成モデルがレビュー内の属性に言及した返信を生成できるようにする。属性語は訓練データから事前に抽出する。全ての単語の全てのレビューにおける TF-IDF を算出し、その TF-IDF 値の上位 1000 件の平均を単語のスコアとし、そのスコアの上位 500 件の単語を属性語とする。2つ目は、定型文のフィルタリングである。定型的な表現で書かれた返信文を訓練データから削除する。これにより、モデルが紋切り型の返信を生成することを抑制することが期待できる。処理の手順は次の通りである。訓練データの返信を文に分割し、それぞれの返信文に対して定型度スコアを算出する。定型度スコアは、返信文における単語の tri-gram の訓練データ全体における出現頻度の平均値として算出する。スコアが最も高い上位 30%の文を定型文とみなし、これらを訓練データから削除する。

返信の統合では、文毎に生成された返信をマージする。この際、重複する返信文を除外する。生成された返信文間の類似度を正規化された編集距離で測定し、この値が 0.1 以下の場合、元のレビューにおける出現順序が後の返信文を保持し、他方を削除する。ただし、属性語を含む返信については除外しない。この手続きの後に残された返信文を、元のレビューと同じ順序で連結し、最終的な返信とする。

提案手法の評価実験では、苦情判定モデルと返信生成モデルをそれぞれ評価した。苦情判定モデルの評価では、テストデータにおけるレビュー文が苦情か否かを分類し、その正解率、精度、再現率、F 値を測った。その結果、訓練時の epoch 数を 1 に設定したときの結果が一番良く、そのときの正解率は 0.8877、精度は 0.8718、再現率は 0.9091、F 値は 0.8901 であった。これにより苦情判定モデルの性能が十分に高いことを確認した。

返信生成モデルの評価では、訓練データに対するフィルタリングの有無や、文毎に返信を生成するかどうかなどの条件を変えた手法を比較した。また、訓練データに対するフィルタリングを行わずに返信生成モデルを学習する手法をベースラインとした。まず、自動評価を行った。生成された返信の品質を BLEU-4 と DISTINCT-4 で評価した。BLEU-4 は提案手法よりもベースラインの方が高かった。これは、ベースラインは定型文をよく生成するが、正解の返信にも定型文が多く、単語 n-gram の一致率が高くなったためと推測される。一方、DISTINCT-4 は訓練データのフィルタリングを行った提案手法で高くなり、定型的な表現の生成の抑制にフィルタリングが有効であることを示した。次に、人手による評価を行った。返信の流暢性、非冗長性、総合評価、属性言及率(レビューに出現する属性のうち返信で言及されているものの割合)を 7 名の被験者が評価した。その結果、属性に言及しない返信のフィルタリングを行う提案手法は、ベースラインに比べて属性言及率が高くなる一方で、流暢性と非冗長性は低くなった。これは、属性言及率と流暢性・非冗長性がトレードオフの関係にあることを示唆する。また、定型文のフィルタリングに焦点を当てると、非冗長性が改善されるとともに、流暢性や総合評価も高くなり、紋切り型の表現の生成が抑制されていることが確認された。さらに、複数の属性を持つレビューに対しては、文毎に返信を生成する手法が、そうでな

い手法に比べて属性言及率と総合評価で優れており、その有効性が確認された。