

レビュー内容の網羅性と緻密性に基づいた有用性判定法の検討

○岡本章平 蒲原智也 波多野賢治 (同志社大学 文化情報学部)



1. 研究背景

■ 商品やサービスのレビュー



効率よく商品の評判に関する情報を取得すること必要

■ 有用なレビューを効率よく取得するための解決策の例

参考になるレビューに投票する機能をつけたウェブサイトも存在

【投票機能の課題】

- 投票するユーザが少ない商品やサービスに対しては効果的でない
- 投稿されて間もないレビューに対しては効果的でない

図1: Amazon[1]のレビュー投票機能

■ 本研究の目的

- 効率よく商品の評判情報を取得するためには、レビューの有用性を判定する必要がある
- レビューは、トピックについて網羅的に述べられているか、あるトピックについて詳しく述べられていることが重要との指摘がある[2]

【研究目的】 網羅性と緻密性の推定手法の提案を行い それらを用いてレビューの有用性判定を行う

[1] Amazon: <https://www.amazon.co.jp/> (2017/11/28 閲覧)
 [2] 田中成典, 北野光一, 寺口敏生, 今城彰子, 柳田尚明, 「広告の特徴に基づく口コミの分類に関する研究」, 情報処理学会論文誌, データベース, Vol. 4, No. 3, pp.22-32, 2011

■ 特徴語を用いた網羅性と緻密性の算出方法

- レビューに記述されている名詞と特徴語との類似度をword2vecとコサイン類似度を用いて算出
- 特徴語と類似度が閾値以上の名詞が属する評価項目の有無を判定 (評価実験で最も精度が高くなった閾値を採用)
- 各トピックの評価項目の有無の合計を算出

表1: レビュー内容と評価項目の一致の有無

【網羅性と緻密性の算出の例】

朝食は十分な量がありました。ありがとうございました。

網羅性: 部屋, 食事, その他のトピックの中で食事のトピックのみ述べられている

緻密性: 食事のトピックについて1点だけ緻密に述べられている

| トピック | 評価項目 | 有無 | 有の合計 |
|------|----------|----|------|
| 部屋 | 通信設備について | 無 | 0 |
| | エアコンについて | 無 | |
| | お風呂について | 無 | |
| 食事 | 飲み物について | 無 | 1 |
| | 朝食について | 有 | |
| ⋮ | ⋮ | ⋮ | ⋮ |

トピックの集合をI, 評価項目の集合をVとする

【レビューxにおける網羅性の度合いを表すCxの算出方法】

N_i : 全トピック数
 n_{ix} : レビューxにおける有の合計が1以上のトピック数

$$C_x = \frac{n_{ix}}{N_i}$$

【レビューxにおける緻密性の度合いを表すDxの算出方法】

n_{ix}^i : レビューxにおける評価項目i番目の有の数 (有: 1, 無: 0とする)

$$D_x = \sum_{i \in V} n_{ix}^i$$

表2: トピックと評価項目

| I | V |
|---------|-------|
| 部屋 | 評価項目1 |
| | 評価項目2 |
| | ⋮ |
| 受付とサービス | 評価項目i |
| | ⋮ |
| ⋮ | ⋮ |
| ⋮ | 評価項目n |

4. 評価実験の方法

提案した緻密性と網羅性を用いて、レビューの有用性が正しく判定できているか検証を行う

- アンケート結果からレビューの有用性のラベル付け

m_{rx} : レビューxにおける参考になったと回答した人数
 Mx : レビューxにおけるアンケート回答者数

$$\text{レビューの有用性の度合い} = \frac{m_{rx}}{Mx}$$

- 目的変数が離散値であるため、交差検証法による多項ロジスティック回帰分析を用いて判定精度を検証 (学習用データ: 400件, テスト用データ: 50件)

2. 関連研究

■ 網羅性の推定手法[3]

形態素解析器MeCab[4]によって得られる以下の品詞情報を特徴語と定義

名詞(一般)
 名詞(ナイ形容詞語幹)
 名詞(非自立)
 接頭詞, 名詞接続

特徴語の保有率によって網羅性を算出

■ 緻密性の推定手法[5]

word2vec[6]により求めた各単語ベクトルの和から各文章の特徴ベクトルを算出

[文1の特徴ベクトル]
 [文2の特徴ベクトル]
 ⋮
 [文nの特徴ベクトル]

隣接する文同士のコサイン類似度が閾値以上となる文同士が連続する回数をもとにその文のトピックの緻密性を算出

【関連研究の問題点】

- レビューに記述されているサービスや商品の特徴について考慮されていない
- 網羅性と緻密性の両方を考慮した提案がなされていない

[3] 別所雅秀, 湯本高行, 新居学, 佐藤邦弘, 「商品レビューの網羅性と支持度の関係の分析」, 研究報告データベースシステム Vol. 2013-DBS-158, No. 19, pp.16, 2013
 [4] Kudo Taku, Yamamoto Kaoru, Matsumoto Yuji, "Applying Conditional Random Fields to Japanese Morphological Analysis", Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.230-237, 2004
 [5] 高島博康, 青野雅樹, 「化粧品レビューサイトにおける口コミの有用性判定」, 言語処理学会第23回年次大会発表論文集, pp.799-802, 2017
 [6] Mikolov Tomas, Chen Kai, Corrado Greg, Dean Jeffrey, "Efficient estimation of word representations in vector space", Proceedings of Workshop at the International Conference on Learning Representations, pp.1-12, 2013

3. 提案手法

■ 宿泊施設の特徴

レビューに記述されている話題は、サービスや商品によって異なるため、サービスや商品ごとの特徴を考慮する必要があると考える

宿泊施設の特徴を表すものとして世界で広く使用されている宿泊施設の評価項目を採用する

宿泊施設の評価項目に記述されている内容を特徴語とする

■ 網羅性と緻密性の定義

【考え方】

【網羅性】
 宿泊施設に述べられるべき項目は、世界で広く使用されている宿泊施設の評価項目にすべて存在すると考える

【定義】

特徴語についてどの程度広く述べられているかを網羅性と定義する

【緻密性】
 ある話題について述べられるべき項目は、世界で広く使用されている宿泊施設の評価項目にすべて存在すると考える

特徴語についてどの程度詳しく述べられているかを緻密性と定義する

■ 使用するデータ

楽天データセットの楽天トラベルの宿泊施設レビュー文620万件から無作為に抽出した450件を使用

■ 正解データの作成

- レビューの有用性のラベル付けを行うために、クラウドソーシングを用いてレビューが参考になるか、参考にならないか二択によるアンケートを実施
- 一件のレビューに対して15人が回答

■ 特徴語の抽出方法

Hotelstars.EU[7]の評価観点を日本語訳したものから形態素解析を行い、そこから抽出した名詞を特徴語とする

[7] Hotelstars.EU: <https://www.hotelstars.eu> (2017/11/28 閲覧)

5. 今後の課題

- 類似度の閾値の設定
- 提案した緻密性と網羅性のレビューの有用性判定の評価実験
- 既存手法と提案手法による有用性判定の比較
- 網羅性と緻密性の推定精度の評価