

旅館レビューからの周辺施設の評判情報の抽出

Extracting information on the reputation of nearby facilities from inn reviews

中村 航太, 溝畑 知風, 延澤 志保 (東京都市大学)

Kouta Nakamura, Chikaze Mizohata, Shiho Hoshi Nobesawa Tokyo City University

P05

研究背景・研究目的・関連研究

研究背景

- ・楽天トラベル¹やじゃらんnet²などの宿泊施設予約サイトには旅館の利用者によるレビューが投稿される。
- ・レビューには旅館内のサービスだけでなく、コンビニを中心とした周辺施設の評判情報も投稿される。

総合★★★★☆ 4 2人中1の方が参考になったと投票しています。

投稿者さんのクチコミ

投稿者さん 2022-05-08 06:53:17

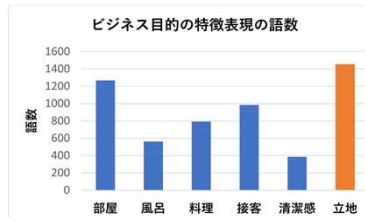
自然に囲まれた旅館だったので、静かに落ち着いて過ごせました。しかし、コンビニが遠い。旅館から最寄駅まで行くには1時間に一本しかない阪急バスを利用、シャトルバスは予約しなければ利用出来ないなど、少し不便さを感じる点がありました。自家用車があればその点スムーズだったと思いますが、大層なため旅館以外の駐車場を利用すれば、駐車料金が高い可能性があります。少し離れたところにはパーキング場やゴルフ場があり、GW中だった事もあり、交通渋滞が至る箇所であり、旅館到着まで少し大変でした。

図1：楽天トラベルのレビュー例

¹楽天トラベル, <https://travel.rakuten.co.jp>. ²じゃらんnet, <https://www.jalan.net>.

関連研究

・宿泊施設のレビューから注目情報を抽出する研究[1]



・旅行目的がビジネス目的の時に立地に関する記述が多い。

・コンビニや最寄り駅に関する記述が多い。

図2：ビジネス目的の特徴表現の語数

立地(周辺施設)に関する情報を重要視する人が多い。

・宿泊施設に投稿されるレビューの分析の研究[2]

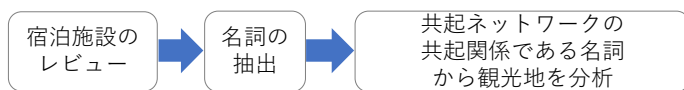


図3：観光地分析の処理の流れ

立地(周辺施設)に関する情報がない。

研究目的

- ・楽天トラベルのレビューの分析から、周辺施設に関する評判情報を抽出する。
- ・旅館レビューから周辺施設について記述されているレビューの抽出と評判情報を抽出する手法を提案する。

提案手法・実験結果

提案手法

入力：楽天トラベルレビュー群

①コンビニに関するレビューの抽出

レビュー例

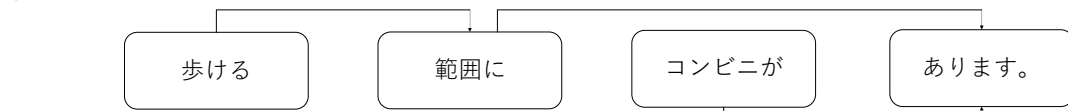
松本駅からのバスはおよそ1時間に一本ですが、湯坂のバス停からすぐです。~~~~~(省略)~~~~~すぐ隣には外湯がありましたが同じ源泉だそうです。歩ける範囲にコンビニがあります。レトロな気分を味わい、静かに過ごせて満足でした。

Wikipedia³に記載されている日本のコンビニエンスストアチェーン一覧の以下の単語がレビューに出現する場合にコンビニに関する記述であると判定。

- ・全国展開するコンビニエンスストア (名称, 略称)
- ・広域に展開するコンビニエンスストア (名称, 略称)
- ・地域限定型展開するコンビニエンスストア (名称, 略称)
- ・鉄道会社系列の全てのコンビニエンスストア (名称, 略称)
- ・コンビニ
- ・コンビニエンスストア

①コンビニに関するレビューの抽出

②CaboCha⁴を用いた係り受け解析



歩ける範囲にコンビニがあります。

図5：係り受け解析による評判情報の抽出例

③コンビニ以外の情報がある

・「コンビニ」の前後に着目し、「や」、「と」、「、」がある場合でこれらの前後に名詞がある場合はその他の周辺施設の情報も抽出されてしまっていると判定する。

余分な情報が含まれる例：徒歩圏内に飲食店、コンビニ、イオンもある。

飲食店 〇 コンビニ 〇 イオン
名詞 名詞

今回はコンビニに関する情報を抽出したいため、以下を出力とする。

1. 徒歩圏内に飲食店がある。
2. 徒歩圏内にコンビニがある。
3. 徒歩圏内にイオンがある。

³Wikipedia, <https://ja.wikipedia.org/wiki/> ⁴CaboCha, <https://taku910.github.io/cabocho/>

図4：評判情報の抽出の処理

図6：係り受け解析によって得られた文の評判情報の抽出の処理

実験結果

コンビニに関するレビューの抽出, 評判情報の抽出の結果 楽天トラベルレビューデータ2019年60レビューを使用。

表1：コンビニに関するレビューの抽出結果

	提案手法		
	コンビニに関する記述ありレビュー	コンビニに関する記述なしレビュー	合計
正解データ	33	0	33
	2	25	27
合計	35	25	60

表2：CaboChaを用いたコンビニのみの評判情報の抽出結果

	CaboCha		
	コンビニのみの情報	コンビニ以外の情報も含む	合計
正解データ	18	14	32
	0	2	2
合計	18	16	34

表3：提案手法を用いたコンビニのみの評判情報の抽出結果

	提案手法		
	コンビニのみの情報	コンビニ以外の情報も含む	合計
正解データ	31	1	32
	0	2	2
合計	31	3	34

・正解データは人手によるレビューからの評判情報の抽出結果とする。

・表2より、CaboChaを用いたコンビニのみの評判情報の抽出は約56%であった。

・表3より、提案手法を用いたコンビニのみの評判情報の抽出は約96%であった。

抽出できなかった文の特徴として以下が挙げられる。

- ・誤字によって係り受け解析が正しく行われない文
- ・コンビニという言葉は含まれるがコンビニの情報が記載されていない文
- ・提案手法の処理では文を分割できなかった文

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様からご提供いただいた「楽天データセット」(https://rit.rakuten.com/data_release/)を利用しました。心より感謝いたします。

参考文献

- [1] 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,” 情報処理学会デジタルブライクティス, Vol.3, No.4, pp.289-296, 2012.
- [2] 落合 夏美, 川合 康史, “宿泊施設レビュー分析による観光地分析,” 情報システム学会第16回全国大会・研究発表大会, p.55-7, pp.1-4, 2020.
- [3] 楽天グループ株式会社 (2020): 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>