S01

テキストマイニングから見る宿泊施設の評価分析

東京理科大学　櫻庭里那　朝日弓未

研究背景

群馬県の人口減少・高齢化問題
・労働人口の減少、街の過疎化
→県の経済、成長力の低下
群馬県の観光業、「温泉」に注目し観光客を増やすことで経済を刺激

研究目的

群馬県の温泉地周辺の宿泊施設をテキストマイニングし、宿泊者の評価基準を知り、効果的な宣伝方法を見つける

利用データ

データ概要：楽天トラベルの施設評価及びレビューデータ
対象：草津、伊香保、水上、四万、磯部温泉地域の宿泊施設
期間：2016～2019年

評価1（立地）	4
評価2（部屋）	4
評価3（食事）	4
評価4（風呂）	4
評価5（サービス	4
評価6（設備）	4
評価7（総合	4
ユーザ投稿本文	初めて利用しました。朝食が充実していてよかったです。空気清浄機があればもっとよかったです。

提案手法(Text Mining Studio)

①単語頻度解析を実施
　レビュー文中の出現頻度の多い語を順に表示
②係り受け頻度解析を実施
　出現頻度の多い語と語の組み合わせを順に表示
　→施設へのイメージを知る

③特徴語分析を実施
　総合評価に１と５をつけた人の口コミを抜き出し比較
　→施設の良い点、悪い点を単語から推測

④評価抽出
　良いイメージ、悪いイメージで語られる言葉を抽出
原文での用いられ方も確認

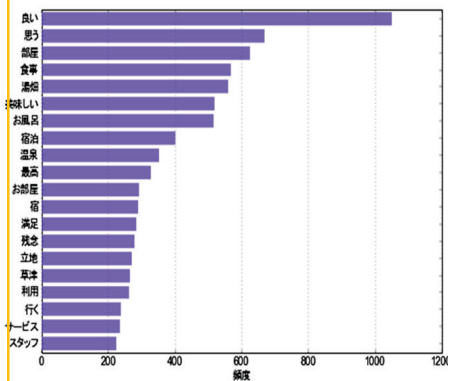

図1　単語頻度解析

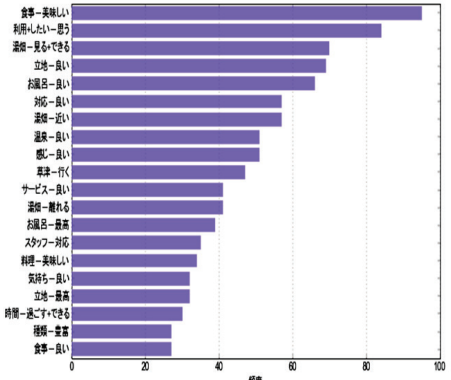

図2　係り受け頻度解析

「美味しい」「湯畑」が上位を占める
→料理の美味しさや湯畑は評価5をつける大きな要因となる

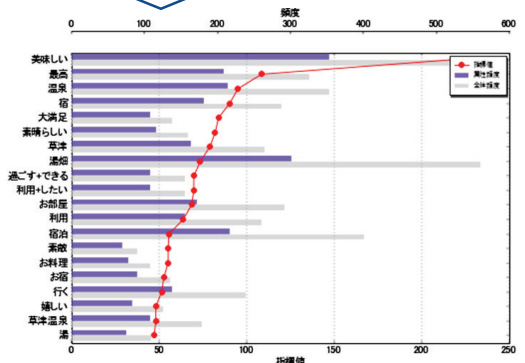

図3　特徴語分析

ネガティブ表現では「部屋」が多い
原文：部屋に子供用のスリッパや歯ブラシなどはなく、その都度フロントへ取りに行って頼んでいました

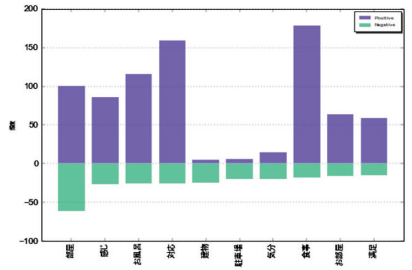

図4評判分析

今後の課題

・各施設のデータ量に差がある
・総合評価１の口コミが少ない

謝辞
本研究の一部は、2021年国立情報学研究所公募型共同研究の助成を受けたものである。ここに記して謝意を表す。