

口コミから抽出した評価観点の重視度に基づく 宿泊施設推薦システムの構築

高松耕太[†] 奥野拓[†]
(公立はこだて未来大学)[†]

1 はじめに

旅行者が宿泊施設を探す際、旅行先や旅行目的、同伴者によって宿泊施設に求める条件が異なることが考えられる。宿泊施設の特徴を調べる方法の一つに口コミがある。しかし、数多くの口コミからユーザの求める情報が記載されている口コミを探すこと、口コミを読むこと自体がユーザにとって負担となっている。

そこで本研究では、口コミから評価観点ごとに評価値を抽出し、ユーザが入力した評価観点の重視度に合った宿泊施設を推薦するシステムを構築する。

2 評価値の抽出手法に関する研究

評価値は、評価内容が肯定的な意味なのか、否定的な意味なのかに応じた値にする必要がある。そのために、肯定的な意味なのか否定的な意味なのかを定量的に表した値である、感情極性値を用いる。感情極性値を算出する際、多くの表現に対応した手法が必要である。

酒井らは、評価情報と種表現との共起文章数として検索ヒット数を用いた、感情極性値の算出手法を提案している [1]。酒井らの手法では、評価情報を「〈対象〉の〈属性〉が〈評価表現〉」のような組み合わせとしている。また、「いい」、「好き」、「最高」を種表現の肯定語、「悪い」、「嫌い」、「最低」を種表現の否定語としている。共起文章数は、評価情報と種表現との共起文章数に加え、評価表現を反対語にした評価情報と種表現との共起文章数を算出することで、感情極性値の精度向上を行っている。この手法では、評価情報を（対象、属性、評価表現）としているため、対象となる口コミに限られてくる。そのため本研究では、酒井らの手法を参考に評価情報を（対象、評価表現）として、感情極性値を算出する。

3 宿泊施設の推薦手法

本研究で構築する、口コミを用いた宿泊施設の推薦手法について述べる。まず、口コミから評価基準となる評価観点を抽出する。さらに、口コミから投稿者の評価として、抽出した評価観点ごとに評価値を抽出する。その後、ユーザが入力した旅行先、旅行目的、同伴者と一致する口コミに絞り込み、宿泊施設の評価を抽出する。抽

出した宿泊施設の評価とユーザの入力した評価観点の重視度との類似度を求め、宿泊施設を推薦する。以降の節では、それぞれの手法について詳しく述べる。

3.1 評価観点の抽出

本研究では、風呂や朝食などの評価基準を評価観点とする。評価観点は、口コミによって異なるが、宿泊施設同士を比較する上で共通の評価観点が必要となる。そこで、使用する全ての口コミデータから網羅的に評価観点を抽出するために、Probabilistic Latent Semantic Analysis (以下 PLSA とする) のモデルを作成する。PLSA とは、文章中から単語と各単語の所属確率をトピックとして抽出し、各トピックの出現率を抽出する手法である。口コミから評価観点となる名詞を抽出し、抽出した名詞を用いてモデルを作成する。その後、作成したモデルのトピックを評価観点とする。作成したモデルのトピックには、抽出した名詞と抽出した名詞の所属確率が含まれている。

3.2 投稿者の評価抽出

本研究では、口コミから評価観点ごとに評価値を抽出したものを投稿者の評価とする。評価値は、-100 から 100 で表し、-100 に近いほど低い評価、100 に近いほど高い評価とする。投稿者の評価を抽出するために、係り受け解析を用いて口コミから（対象、評価表現）となる評価情報を抽出する。「露天風呂」と「ある」などの名詞と動詞「部屋」と「きれい」などの名詞と形容詞の組み合わせとなる係り受け表現を抽出し、名詞を評価情報の対象、動詞と形容詞を評価情報の評価表現とする。抽出した評価情報と 3.1 節で述べた PLSA のモデルを用いて、各評価観点の感情極性値を算出する。まず、評価情報と種表現との共起文章数の和を算出する。種表現は、酒井らの手法と同様の種表現を用いる。また、共起文章数も酒井らの手法と同様に検索ヒット数から算出する。共起文章数の和は、酒井らの手法を参考に評価情報を（対象、評価表現）と置き換えた式 (1) により算出する。

$$\sum_i hits(\text{対象 評価表現 種表現}) \quad (1)$$

評価情報と種表現の肯定語との共起文章数の和を $hits(pos)$ 、否定語との共起文章数の和を $hits(neg)$ で表す。さらに、評価表現を反対語に置き換えた評価情報と肯定語との共起文章数の和を $hits(antpos)$ 、否定語との共起文章数の和を $hits(antneg)$ で表す。共起文章数の算出後、酒井らの手法と同様に式 (2) により感情極性を算出する。

Development of hotels recommender system based on the degree of importance of evaluation perspectives extracting from the review of hotels

[†] Kouta Takamatsu, Taku Okuno

[†] Future University Hakodate

表 1 投稿者の評価抽出例

宿泊施設	旅行目的	同伴者	部屋	サービス	利便性	食事	...	風呂
宿泊施設 1	レジャー	家族	75	0	30	25	...	40
宿泊施設 2	ビジネス	仕事仲間	20	100	10	30	...	20
宿泊施設 2	ビジネス	一人	-20	0	100	30	...	0
宿泊施設 1	レジャー	一人	25	25	-10	15	...	5
...
宿泊施設 50	レジャー	恋人	20	50	50	40	...	50

表 2 宿泊施設の特徴抽出例

宿泊施設	部屋	サービス	利便性	食事	...	風呂
宿泊施設 1	25	12.5	-20	20	...	22.5
宿泊施設 2	0	50	45	30	...	10
宿泊施設 3	30	10	20	30	...	50
宿泊施設 4	20	0	10	60	...	0
...
宿泊施設 50	20	20	20	20	...	20

$$\frac{hits(pos)}{hits(pos) + hits(neg)} - \frac{hits(antpos)}{hits(antpos) + hits(antneg)} \quad (2)$$

本研究における感情極性値は、-100 から 100 で表し、100 に近いほど肯定的な表現、-100 に近いほど否定的な表現とする。算出した感情極性値から各評価観点の平均値を算出し、算出した平均値を投稿者の評価とする。評価されていない評価観点については、評価値を 0 とする。さらに、口コミには、宿泊施設名と旅行目的、同伴者が付与されているため、これらも投稿者の評価に付与する。投稿者の評価抽出例を表 1 に示す。

3.3 ユーザの入力項目

ユーザは、評価観点ごとに重視度を 0 から 100 で入力する。重視度は、重視しているほど 100 に近い値とする。さらに、推薦対象とする宿泊施設を絞り込むために旅行先を選択し、宿泊施設の評価の抽出対象とする投稿者の評価を絞り込むために旅行目的と同伴者をそれぞれ選択する。これらは、投稿者の評価に付与されているため、それぞれの選択肢から選択する。

3.4 宿泊施設の特徴抽出

本研究では、投稿者の評価と同様に評価観点ごとに評価値を -100 から 100 で表したものを宿泊施設の評価とする。まず、3.3 節で述べた旅行先と旅行目的、同伴者と一致する投稿者の評価に絞り込む。絞り込んだ投稿者の評価から同じ宿泊施設ごとに投稿者の評価を集約する。そのために、各評価観点における評価値の平均値を宿泊施設ごとに算出する。表 2 のように算出した評価観点の平均値を宿泊施設の評価とする。

3.5 宿泊施設の推薦

3.3 節で述べた評価観点の重視度と 3.4 節で述べた宿泊施設の評価との類似度の高い宿泊施設を推薦する。そのためにコサイン類似度を用いて類似度の算出を行う。算出した類似度を用いて、類似度の高い順に宿泊施設を推薦する。

表 3 抽出したトピックの上位 5 つの名詞と所属確率

TOPIC 1	デニッシュ	市場	通り	途中	ラーメン
	0.13007	0.07116	0.07116	0.07115	0.07109
TOPIC 2	感動	場	態度	挨拶	アップ
	0.15783	0.15747	0.08606	0.08599	0.01510
TOPIC 3	奇跡	ところ	インテリア	書等	設定
	0.11651	0.11650	0.06363	0.06356	0.06354
TOPIC 4	タイプ	問題	温泉	リピーター	週末
	0.16940	0.11645	0.06359	0.06357	0.06319
TOPIC 5	北海道	騒が	観光	責任	感じ
	0.11897	0.08181	0.08180	0.08179	0.08177

4 抽出可能な評価観点の検証

実際の口コミを用いて、3.1 節で述べた PLSA のモデルを作成し、どのような評価観点が抽出されるのか検証を行った。検証には、楽天トラベルの口コミから Hotel & Spa Resort LA VISTA 函館ベイの 2018 年 1 月から 12 月までの口コミ 145 件を用いた。今回は、Google が公開している Cloud Natural Language API の形態素解析ツールである Syntax を用いて、名詞を抽出し、PLSA のモデルを作成した。モデルを作成する際、主要な評価観点を 5 つと仮定し、トピック数を 5 と設定した。検証の結果から、各トピックから所属確率の高い 5 つの名詞と所属確率を表 3 に示す。検証の結果、意味ごとの分類ができていないため、各トピックが何の評価観点を表しているのか把握できなかった。

5 評価観点の提示手法

本研究では、抽出したトピックを評価観点とするが、意味ごとに単語を分類するには限界があると考えられる。そのため、抽出したトピックをユーザへ提示する手法を検討する必要がある。嘉村らは、トピックを客観的に解釈可能となるトピック表現語を抽出する手法を提案している。嘉村らの評価結果では、特定単語の所属確率を全トピックにおける特定単語の所属確率の和に対する割合として算出し、一番値の大きい単語をトピック表現語とする手法が最も良い結果となっている。そのため本研究でも、嘉村らの手法を用いて評価観点の提示を目指す。

6 まとめ

本稿では、ユーザの宿泊施設の支援を目的とした、口コミを用いた宿泊施設推薦システムについて述べた。今後は、5 章で述べた手法についての調査および検証を行う。その後、システムの構築および評価を行う。

参考文献

- [1] 酒井義和 荒木健治：反対語を利用した文脈依存評価表現の感情極性判定，電子情報通信学会論文誌，pp.1778-1789 (2010).
- [2] 嘉村巨太 黄宏軒 川越恭二：トピックモデルにおけるトピック表現語導出，DEIM Forum 2011 C1-4 (2011).