

1. はじめに
4. まとめ

[論文紹介] ホテルの評価値予測のための カテゴリ特化型感情極性辞書

水上 和秀 (Kazuhide Mizukani)
u355020@st.pu-toyama.ac.jp

富山県立大学 工学部 電子情報工学専攻

February 20, 2024

1 はじめに

2/18

1. はじめに
4. まとめ

背景

ホテル予約サイトでホテルの宿泊客から投稿された各ホテルにタイル評価が提供されるが、一部の評価値がレビュー内容と矛盾しているものが存在し、そのような評価はサイト利用者を混乱させる可能性がある。

目的

レビュー文章から各カテゴリに対する評価値を予測する手法を提案する。

提案手法

3/18

提案の流れ

- 1. はじめに
- 4. まとめ

- 1 レビューデータセット R の各レビューを分ごとに区切り、各分で使用される単語から、その分に記述されているカテゴリのラベルを文に割り当てる。
- 2 レビューごとに、各カテゴリを説明する部分のみを抜粋し、各カテゴリの各評価値に分類する
- 3 分類されたレビューと各単語の出現率に基づいて、カテゴリごとに感情極性地を計算し、カテゴリ特化型感情極性辞書を作成する
- 4 作成したカテゴリ特化型感情極性辞書を使用して、カテゴリに分類された分の単語に感情極性値を割り当てることで文の特徴量を計算し、評価値を予測する

カテゴリ辞書

- 文章に記述されているカテゴリのラベルをその分に割り当てるために、各カテゴリに関する単語をまとめたカテゴリ辞書を作成する。
- カテゴリ辞書は、田熊らが作成した辞書に基づいている。これは筑波大学文単位評価極性タグつきコーパスで使用されている上位 500 の名詞を人手でカテゴリに分類したものである。
- 彼らの辞書から、単語がレビューに含まれている場合にそのレビューに記述されているカテゴリを特定できるような 315 語のみ人手で抽出した。

サービス	立地	部屋	設備	風呂	食事
愛想	駅	TV	ベッド	温泉	デザート
価格	観光	ベッド	階段	浴槽	魚
:	:	:	:	:	:

Table 1: カテゴリ辞書の例

カテゴリリレビューの分類

5/18

1. はじめに
4. まとめ

- データセット R の各レビューを「。」、「.」、「！」、「？」など文の終端を表す記号に基づいて文に分割する。
- カテゴリ辞書を用いて分割した分をカテゴリに分割し、カテゴリラベルを割り当てる。
- レビューを $r \in R$ 、カテゴリラベルを $X \in \{\text{サービス}, \text{立地}, \text{部屋}, \text{設備}, \text{風呂}, \text{食事}\}$ とし X に関する文の集合 r_X をレビュー r ごとに作成するまた、 R_X をすべてのレビューから抽出した X のカテゴリレビューの集合とする
- 次に、各カテゴリレビュー r_X を X に対する評価値 $Y \in \{1, 2, 3, 4, 5, *\}$ (*は評価なし) と分類する。つまり、r に付属している X に対する評価値が Y であった場合に、 r_X をカテゴリレビュー集合 R_{XY} に分類する

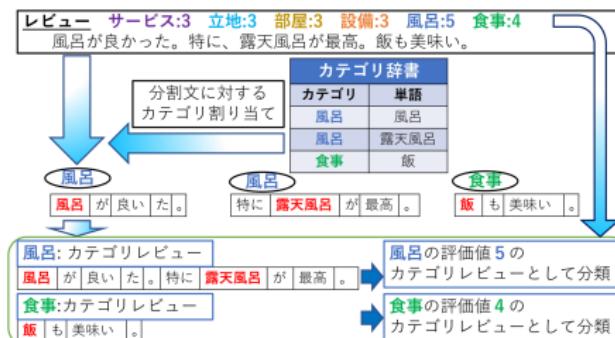


Figure 1: カテゴリレビュー分類の例

感情極性辞書の作成 1

6/18

1. はじめに
4. まとめ

カテゴリリレビューからカテゴリ特化型感情極性辞書を作成する。感情極性に使用する単語は「形容詞」、「形容動詞」、「副詞」、「名詞」の 4 つの品詞とする。データセット R においていずれかのカテゴリ X の R_X に含まれるこれら 4 つの品詞の単語は感情極性語となる。

各感情極性語 A のカテゴリ X における感情極性値 E_{XA} を計算する方法は R_{XY} における A の出現率に基づいている。

1. まず、以下の式を用いて X におけるすべての評価値での感情極性語 A の出現率 U_{XA} を計算する

$$U_{XA} = \frac{\sum_{Y=1}^5 |R_{XYA}|}{\sum_{Y=1}^5 |R_{XY}|} \quad (1)$$

2. R_X 内すべての感情極性語について X 出の出現率を計算した後、以下の式でそれらの平均値 $U_{X_{ave}}$ を計算する。 $(N$ は感情極性語の語彙数)

$$U_{X_{ave}} = \frac{\sum_{k=1}^N U_{X_k}}{N} \quad (2)$$

感情極性辞書の作成 2

7/18

1. はじめに
4. まとめ

3. X での感情極性語 A の感情極性スコア S_{X_A} を、つぎの式で計算する

$$S_{X_A} = \frac{\sum_{Y=1}^5 Y(U_{XYA} + U_{X_{ave}})}{(U_{XYA} + U_{X_{ave}})} \quad (3)$$

4. R_X 内のすべての感情極性語の X における感情極性スコアを計算した後、次の式で E_{X_A} を計算する

$$E_{X_A} = \frac{S_{X_A} - S_{X_{ave}}}{\sigma_X} \quad (4)$$

ここで、 $S_{X_{ave}}$ は、 X におけるすべての感情極性語の感情極性スコアの平均であり、 σ_X は X におけるすべての感情極性語の感情極性スコアの標準偏差である。

カテゴリ X	語彙数	$U_{X_{ave}} \times 10^5$	$S_{X_{ave}}$	σ_X	最大/最小 E_{X_A}		評価値 Y の出現率 $U_{XYA} \times 10^5$					感情極性	
					単語 A	親切	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	S_{X_A}	E_{X_A}
サービス	3,849	74	2.9359	0.2054	親切	1,283	1,923	2,265	5,098	9,401	3.95429	4.95696	
					二度と	1,710	365	84	14	12	1.53484	-6.81945	
立地	3,173	78	2.9662	0.1710	抜群	50	53	56	163	585	3.90786	5.50419	
					悪い	9,068	3,547	2,106	1,134	577	1.84725	-6.54115	
部屋	3,866	80	2.9561	0.1817	広々	153	217	322	907	1,940	4.08228	6.19599	
					汚い	5,847	1,815	561	223	99	1.53723	-7.80674	
設備	3,574	79	2.9688	0.1638	豊富	69	155	212	420	1,023	3.95444	6.01701	
					汚い	3,434	1,207	638	233	130	1.74419	-7.47544	
風呂	3,013	81	2.9740	0.1873	素晴らしい	182	173	278	649	2,246	4.16928	6.37931	
					汚い	3,960	1,289	504	200	73	1.62283	-7.21080	
食事	3,242	79	2.9487	0.1976	豊富	788	1,159	1,699	4,569	7,715	4.05749	5.60982	
					ひどい	2,057	296	103	49	20	1.52046	-7.22598	

Table 2: 感情極性辞書の数値

感情特徴量の抽出

8/18

1. はじめに
4. まとめ

- 作成したカテゴリ特化型感情極性辞書を使用して各カテゴリレビューから特徴量を抽出する。
- カテゴリレビュー r_X に含まれるすべての感情極性語に、感情極性値を割り当てる。
- カテゴリ特化型感情極性辞書において感情極性語 A のカテゴリ X での感情極性値が E_{XA} であるとすると、カテゴリーレビュー r_x 中で A の次の単語が否定語「ない」の場合、 r_x における A の感情極性値として $E_{XA} \times (-1)$ を使用する。そして、 r_X のすべての感情極性語の感情極性値の平均を計算し、これを r_X の感情特徴量とする。



Figure 2: 感情特徴量抽出の例

数値実験

9/18

データ数

本実験では楽天トラベルデータセットからの 2 種類のデータを用いた。1 つめは 2004 年から 2015 年の 1442695 件のレビューを含むデータである。この一部は評価値予測のモデルを作成するためのトレーニングデータとして用いた。もう一つは 2016 年から 2019 年の 1914418 件のビューを含む新データである。新データの一部は作成したモデルを評価するためのテストデータに用いた。

	評価値	1	2	3	4	5	合計	Sub (%)
旧 データ	サービス	21,164	31,956	162,680	292,848	317,222	825,870	91.93
	立地	2,281	16,718	61,225	217,741	365,760	663,725	83.99
	部屋	21,741	51,687	135,273	29,348	342,933	845,117	80.37
	設備	16,892	44,410	140,493	193,902	157,812	553,509	82.78
	風呂	18,169	36,396	56,030	105,716	152,589	368,900	76.27
	食事	13,702	31,076	71,131	167,821	209,811	493,541	89.04
新 データ	サービス	22,579	30,544	105,579	207,398	330,932	697,032	92.18
	立地	2,406	15,282	60,145	165,426	244,721	487,980	86.80
	部屋	23,552	47,038	111,240	233,472	288,757	704,059	78.21
	設備	17,604	40,753	124,569	162,468	141,545	486,939	85.87
	風呂	17,974	35,023	63,905	158,623	248,493	524,018	77.29
	食事	25,278	44,804	96,177	248,635	422,296	837,190	84.36

Table 3: 各データセットのカテゴリレビューの数

数値実験 1 概要

10/18

1. カテゴリレビューの妥当性

- 被験者へのアンケートによってカテゴリレビューの妥当性を評価した。各カテゴリレビュー r_x がそのカテゴリ X を説明しているか否かを評価する。
- 旧データから各カテゴリのカテゴリレビューをランダムに 20 件ずつ抽出し、各カテゴリレビューに対して 3 人の被験者を割り当てた
- 被験者は与えられたカテゴリレビューを読み、各カテゴリレビューに記載されている内容のカテゴリを選択する

年齢	20 代	30 代	40 代	50 代	60 代以上	合計
男性	1	1	2	2	0	6
女性	1	0	2	2	1	6
合計	2	1	4	4	1	12

Table 4: 実験 1.2 の被験者の内訳

数値実験 1 結果

11/18

1. はじめに
4. まとめ

1. カテゴリレビューの妥当性の結果

- r_X にたいして 3 人の被験者のうち少なくとも 1 人が正解した場合に r_X は X について妥当であるとした。
- 妥当とされたカテゴリレビューの割合は全体の 90.83 % となった。結果より、カテゴリレビューの妥当性は十分に高いことがわかる

正解者数	サービス	立地	部屋	設備	風呂	食事	合計
1 人	1	1	2	3	1	3	11
2 人	4	0	4	5	4	1	18
3 人	13	15	11	11	14	16	80
合計	18	16	17	19	19	20	109

Table 5: 妥当であるとされたカテゴリレビュー数

数値実験 2

12/18

2. レビューから受ける印象と評価値の違い

- 次に、カテゴリリレビューから人の目で評価値を予測した際の実際にレビューによって付加された評価値との差を検証する
- 同様の被験者に自身が選択したカテゴリに対してレビューがどのような評価値をつけているかを予測してもらった。

- 各カテゴリリレビューに対する 3 人の被験者による予測値の平均値を四捨五入した値と、レビューによる評価値の混同行列は以下のようになった。
- 表より、レビューと閲覧者の間でレビューがポジティブな意見を表しているかネガティブな意見を表しているかについては感覚的近くに大きな違いはないことがわかる。

		被験者による予測値の平均				
		1	2	3	4	5
評 価 値	1	13	6	0	1	1
	2	5	11	4	2	0
	3	1	13	4	4	0
	4	0	2	0	8	11
	5	0	1	0	3	19

Table 6: 実験 2 の混同行列

数値実験 3 概要

13/18

3. カテゴリ評価値予測

- カテゴリ特化感情極性辞書の特徴量をもちいて機械学習を行い、レビューのカテゴリ評価値の予測の精度を検証する。
- 分類器として `scikit-learn` のロジスティック回帰を用い、分類期の学習のために、旧データの各カテゴリのカテゴリレビューからランダムに 7 万件を選択し、トレーニングデータとした。
- テストデータは新データから各カテゴリに対してランダムに抽出された 20 万件のカテゴリレビューで構成される
- 感情特徴量のみを特徴量として使用し、各カテゴリの評価値を予測する。比較する辞書は、カテゴリ特化型感情極性辞書 (CSPD)、日本語評価極性辞書 (JSP)、単語感情極性対応表 (PN Table) とする。

日本語評価極性辞書

東北大の乾・岡崎研究室が作成した極性対応表。辞書評価極性を持つ名詞 (句) 8,500 件に対して評価極性情報を人手で付与したもの。

単語感情極性対応表

東京工業大の高村研究室が作成した極性対応表。単語に -1 ~ 1 の感情極性値を付与したもの

数値実験 3 結果

評価値を予測したときの結果は答えの評価値を Y 、予測値を Y_e としたとき

正答率: すべてのカテゴリレビューのうち、 $Y = Y_e$ となった割合

適合率: Y_e となったカテゴリレビューのうち、 $Y = Y_e$ となった割合

再現率: Y のカテゴリレビューのうち、 $Y = Y_e$ となった割合

F1 値: 適合率と再現率の調和平均値、二乗平均誤差

の 5 つの指標で評価する

→結果より、カテゴリ特化型感情極性辞書による感情特徴量はほかに辞書による感情特徴量よりもカテゴリ用か予測において優れた特徴量であることが分かった

		サービス	立地	部屋	設備	風呂	食事	平均
正答率	CSPD	0.4666	0.5055	0.4507	0.3780	0.4932	0.5126	0.4678
	PN Table	0.4727[†]	0.5014 [†]	0.4292 [†]	0.3432 [†]	0.4722 [†]	0.5091 [†]	0.4546 [†]
	JSD	0.4648 [†]	0.5014 [†]	0.4190 [†]	0.3392 [†]	0.4676 [†]	0.5028 [†]	0.4491 [†]
適合率	CSPD	0.2907	0.2409	0.3402	0.2795	0.2766	0.3021	0.2883
	PN Table	0.1596 [†]	0.1003 [†]	0.1564 [†]	0.2149 [†]	0.1384 [†]	0.1784 [†]	0.1580 [†]
	JSD	0.1573 [†]	0.1003 [†]	0.1453 [†]	0.1338 [†]	0.1366 [†]	0.1839 [†]	0.1429 [†]
再現率	CSPD	0.2513	0.2084	0.2594	0.2496	0.2718	0.2692	0.2516
	PN Table	0.2095 [†]	0.2000 [†]	0.2191 [†]	0.2209 [†]	0.2373 [†]	0.2094 [†]	0.2160 [†]
	JSD	0.2088 [†]	0.2000 [†]	0.2089 [†]	0.2111 [†]	0.2453 [†]	0.2148 [†]	0.2148 [†]
F1 値	CSPD	0.2469	0.1552	0.2509	0.2339	0.2466	0.2708	0.2341
	PN Table	0.1660 [†]	0.1336 [†]	0.1765 [†]	0.1948 [†]	0.1724 [†]	0.1619 [†]	0.1675 [†]
	JSD	0.1649 [†]	0.1336 [†]	0.1510 [†]	0.1425 [†]	0.1750 [†]	0.1641 [†]	0.1552 [†]
RMSE	CSPD	1.0297	1.0572	1.0470	1.0663	1.1616	0.9934	1.0592
	PN Table	1.2571 [†]	1.0923 [†]	1.2512 [†]	1.1298 [†]	1.3629 [†]	1.2351 [†]	1.2214 [†]
	JSD	1.2376 [†]	1.0923 [†]	1.3475 [†]	1.0765 [†]	1.3577 [†]	1.2388 [†]	1.2251 [†]

Table 7: 感情極性辞書の感情特徴量のみを使用した評価予測の結果

数値実験 4 概要

15/18

4. カテゴリ評価値予測

- 次に、カテゴリ特化型感情極性辞書を用いて各カテゴリレビューから作成した感情特徴量をそのカテゴリレビューを BERT を用いてベクトル化したものと組み合わせることを考える。
- さらに、既存の感情極性辞書を用いて作成した感情特徴量と BERT によるベクトルを組み合わせた場合と、ベクトル化手法として fastText を用いた場合とも比較する

BERT

- 「BERT」は google が開発した学習済み自然言語モデルであり、単語をベクトル化し単語の特徴量を抽出し言語を分析する

fastText

- 「fastText」は facebook が公開した学習済み自然処理ライブラリで、単語をベクトル化することで単語間の距離を計算しコンピュータ上での言葉の処理を可能にしている

数値実験 4 結果

16/18

比較

手法として

BERT のみを用いて単語をベクトル化

BERT とカテゴリ特化型感情極性辞書を組み合わせてベクトル化

BERT と単語感情極性対応表を組み合わせてベクトル化

BERT と日本語評価極性辞書を組み合わせてベクトル化

fastText のみを用いてベクトル化、

fastText とカテゴリ特化型感情極性辞書を組み合わせて単語をベクトル化

して得られたベクトルを分類器への入力とし、各カテゴリの評価値を予測した。

- 結果より、カテゴリ特化型感情極性辞書の感情特微量と BERT の組み合わせにより、BERT によるベクトルを単独で使用した場合よりも予測精度が高くなつた
- さらに、BERT と組み合わせる感情特微量として、ほかの感情極性辞書によるものを用いた場合と比較するとカテゴリ特化型感情極性辞書による感情特微量を用いたほうがより大きな精度の向上が認められた
- また、文脈依存でないベクトル化手法である fastText を用いた場合でも、カテゴリ特化型感情特性辞書の感情特微量を組み合わせることにより、精度が向上している。
- fastText よりも BERT を用いることで予測精度が向上していることもわかる

数値実験 4 結果

17/18

1. はじめに
4. まとめ

		サービス	立地	部屋	設備	風呂	食事	平均
正答率	BERT	0.5486	0.5368*	0.5032 [†]	0.4309	0.5414*	0.5806 [†]	0.5236 [†]
	BERT+CSPD	0.5493	0.5376	0.5058	0.4318	0.5431	0.5826	0.5250
	BERT+PN	0.5484*	0.5372	0.5036 [†]	0.4308	0.5410*	0.5810*	0.5237 [†]
	BERT+JSD	0.5489	0.5369	0.5036 [†]	0.4307	0.5419	0.5807	0.5238 [†]
	fastText	0.5350 [†]	0.5319 [†]	0.4888 [†]	0.4204 [†]	0.5307 [†]	0.5624 [†]	0.5115 [†]
	fastText+CSPD	0.5370 [†]	0.5333 [†]	0.4918 [†]	0.4221 [†]	0.5324 [†]	0.5656 [†]	0.5137 [†]
適合率	BERT	0.4488*	0.4098	0.4305 [†]	0.4020	0.4192 [†]	0.4546*	0.4275*
	BERT+CSPD	0.4508	0.4162	0.4352	0.4032	0.4234	0.4581	0.4311
	BERT+PN	0.4506	0.4356	0.4331 [†]	0.4019	0.4189 [†]	0.4552	0.4325
	BERT+JSD	0.4490	0.4443 *	0.4330	0.4021	0.4203*	0.4540	0.4338
	fastText	0.4216 [†]	0.4171	0.4006 [†]	0.3740 [†]	0.4132	0.4389 [†]	0.4109 [†]
	fastText+CSPD	0.4253 [†]	0.3502 [†]	0.4048 [†]	0.3798 [†]	0.4134 [†]	0.4434 [†]	0.4028 [†]
再現率	BERT	0.3976 [†]	0.2655 [†]	0.3744*	0.3352	0.3874	0.3964	0.3594 [†]
	BERT+CSPD	0.4023	0.2710	0.3799	0.3360	0.3887	0.4021	0.3633
	BERT+PN	0.3997	0.2666*	0.3734*	0.3345	0.3879*	0.3989	0.3602 [†]
	BERT+JSD	0.4002	0.2643 [†]	0.3735*	0.3362	0.3897	0.3997	0.3606 [†]
	fastText	0.3322 [†]	0.2439 [†]	0.3094 [†]	0.2898 [†]	0.3393 [†]	0.3295 [†]	0.3074 [†]
	fastText+CSPD	0.3404 [†]	0.2464 [†]	0.3156 [†]	0.2929 [†]	0.3445 [†]	0.3418 [†]	0.3136 [†]
F1 値	BERT	0.4115*	0.2693 [†]	0.3895 [†]	0.3477	0.3923 [†]	0.4158*	0.3710 [†]
	BERT+CSPD	0.4153	0.2779	0.3952	0.3488	0.3952	0.4213	0.3756
	BERT+PN	0.4146	0.2710*	0.3898*	0.3473	0.3926 [†]	0.4183	0.3723 [†]
	BERT+JSD	0.4142	0.2674 [†]	0.3899 [†]	0.3490	0.3948	0.4182	0.3723 [†]
	fastText	0.3476 [†]	0.2332 [†]	0.3189 [†]	0.2850 [†]	0.3532 [†]	0.3505 [†]	0.3148 [†]
	fastText+CSPD	0.3539 [†]	0.2372 [†]	0.3265 [†]	0.2897 [†]	0.3579 [†]	0.3645 [†]	0.3216 [†]
RMSE	BERT	0.9362 [†]	0.9059 [†]	0.9766 [†]	1.0309 [†]	0.9976 [†]	0.9039 [†]	0.9585 [†]
	BERT+CSPD	0.9270	0.8999	0.9635	1.0236	0.9878	0.8887	0.9484
	BERT+PN	0.9312*	0.9038*	0.9755 [†]	1.0297*	0.9956 [†]	0.8981 [†]	0.9557 [†]
	BERT+JSD	0.9312 [†]	0.9061 [†]	0.9755 [†]	1.0290*	0.9918	0.8971 [†]	0.9551 [†]
	fastText	1.0214 [†]	0.9340 [†]	1.0278 [†]	1.0532 [†]	1.0308*	0.9802 [†]	1.0079 [†]
	fastText+CSPD	1.0016 [†]	0.9274 [†]	1.0059 [†]	1.0442 [†]	1.0207 [†]	0.9543 [†]	0.9924 [†]

Table 8: カテゴリレビューベクトルと感情特徴量を使用した評価予測の結果

1. はじめに
4. まとめ

おわりに

- ホテルのレビューから各カテゴリに対する評価値を予測する方法を提案した
- 各カテゴリにおける単語の出現率を使用して、各カテゴリに特化した感情極性辞書をさくせいした
- 提案した感情極性辞書は、BERT と組み合わせることで、BERT のみを使用する場合よりも優れた結果を示した