

レビュー文に基づいた星評価項目の細分化・自動スコアリングの提案

背景

レビュー文

投稿者さん
 2名で宿泊しました。
 立地がとてよく、駅近なのが最高でした。
 部屋も可愛らしく、とても快適に過ごせました。
 都内行く時はまた利用したいと思います。

- レビュー内には**多様な情報**が点在
- 内容を素早く読み取ることは**難しい**

星評価

項目別の評価

- サービス ★★★★★ 4.10
- 立地 ★★★★★ 4.51
- 部屋 ★★★★★ 4.16
- 設備・アメニティ ★★★★★ 4.03
- 風呂 ★★★★★ 3.98
- 食事 ★★★★★ 4.37

総合評価
 ★★★★★ 4.30
 アンケート件数：2586件



- 予め用意されている項目のみ
- ユーザーによって評価の**ニュアンスが異なる**可能性

目的

レビュー文から算出した評価値を基準に

- 星評価項目を**細分化**
- その評価項目で**自動スコアリング**

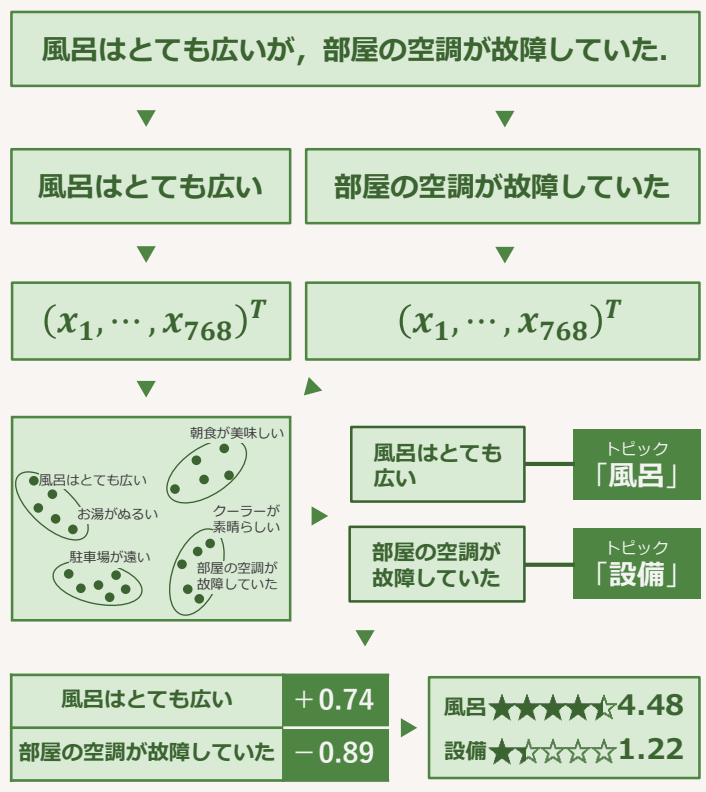
▶ より詳しい情報をレビュー文からの評価値で**素早い読み取りが可能**

使用データ

楽天トラベルレビュー文：アスペクト・センチメントタグ付きコーパス
 文の数：76,624 レビュー文：24,310

提案手法

- レビュー文の分割**
 複数のトピックを含みうるレビューの1文を**接続助詞と係り受け情報**を用いてトピック構造ごとの文節に分割する
- 文節のベクトル化**
 Sentence BERTにより各文節を768次元のベクトルに変換
- クラスタリング**
 文節ベクトルに対して**スペクトラルクラスタリング**を実行し意味内容の近いクラスタに分類
- 感情分析による評価値生成**
 文節に対して**アスペクト感情分析**を行い算出された極性値を星評価値に整形



結果・考察

目的①についてのクラスタ数別の精度
 文節をレビュー文に復元して正解データと比較

手法\クラスタ数※	6	7	8	9	10
BERT	0.723	0.716	0.730	0.760	0.776
Sentence BERT (not fine-tuned)	0.695	0.740	0.752	0.760	0.753
Sentence BERT (fine-tuned)	0.882	0.898	0.890	0.875	0.877

細分化トピック	駅情報		立地
	宴会		
	唐津駅までも散歩がてら歩ける距離内です		
	彦根駅前なので鉄道利用の場合非常に便利です		
正月限定の生ビールサービスはとて良いです			
アルコール飲み放題が1900円なので、もう少し安いと良い			

今後の課題

- 手法 **STEP4**の実装
- より精緻なレビュー分割
- 全体にかかる評価表現の処理

※コーパスとの精度を表示
 7クラスタ以上は内6つ選択の最大値

安部克, & 中島伸介. (2020). レビュー自動スコアリング方式に基づくホテル推薦システム. In *DEIM Forum* (pp. P1-28).
 月岡晋吾, 吉川大弘, & 古橋武. (2015). 係り受け情報を用いたトピック粒度の細分化に関する検討. In *人工知能学会全国大会論文集 第29回 (2015)* (pp. 3L35-3L35). 一般社団法人人工知能学会.