

1.研究背景

- ・ECサイトには商品・サービスが多くある
 - ↳ 膨大なレビューがあり読むのは時間がかかる
 - ↳ ユーザが商品を選別するのが困難である
- ・カメラアプリの比較のためのアプリ分類[1]

アプリ説明文

→ 名詞の抽出 →アプリのクラスタリング

レビュー

→ 特徴語の抽出→アプリ評価の可視化
 - ↳ アプリの機能や特徴を提示した比較の補助を目的
- ・レビュー構造の俯瞰のための評価項目抽出[2]

評価項目の出現頻度 + レビュー中の構文特徴
 - ↳ 対象商品の評価項目の抽出

例) 「デザインが**かっこいい**。」→「デザイン」

「使用感が**良い**です。」→「使用感」

2.研究目的

- ・レビュー中の評価項目を用いた商品のクラスタ分析
- ・末端商品カテゴリで重要とされてる評価項目の判定

3.提案手法

1.化粧水商品 A のレビュー群

抽出ボタン

評価表現辞書[3]を用いた評価項目抽出

肌につけてもサラッとしていて**香り**もとても良いです！

評価表現を含む文節の係り元文節の内の名詞を**評価項目**とする

2. 評価項目の抽出

評価項目	出現頻度
肌	780
香り	650
...	...

3. 商品ベクトルの作成 (正規化)

評価項目	肌	香り	敏感	乾燥	...	柑橘
商品A	0.078	0.065	0.012	0.008	...	0.004
商品B	0.065	0.053	0.001	0.007	...	0.012
商品C	0.072	0.012	0.048	0.039	...	0.000
商品D	0.066	0.007	0.055	0.028	...	0.001

4. クラスタ分類

商品A	0.078	0.065	0.012	0.008	...	0.004
商品B	0.065	0.053	0.001	0.007	...	0.012
商品C	0.072	0.012	0.048	0.039	...	0.000
商品D	0.066	0.007	0.055	0.028	...	0.001

1.特定の商品カテゴリについて各商品に書き込まれたレビューを群として入力する

2.抽出パターンにより評価項目を抽出

3.評価項目と出現頻度の辞書作成, TF値の商品ベクトルを作成

4.クラスタリング手法を用いて評価項目に着目した商品分類を行う

5.商品のクラスタ結果と各クラスタの特徴語を分析する

入力：商品カテゴリの各商品レビュー群

レビュー中の評価項目の抽出

評価項目を用いた商品ベクトルの作成

商品のクラスタ分類

出力：商品分類結果と各クラスタの特徴語

図1:本手法のフローチャート

4.結果

5.考察

使用データ ：楽天市場 商品レビューデータ[4]
商品カテゴリ：化粧水・ローション
商品数 ：50 商品

レビュー数 ：41,087 レビュー
次元数 ：238次元

K-means++法を用いてクラスタ分類(k=5)を行った.
各クラスタでの出現頻度が多い評価項目の上位15単語. (表1)

	商品数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
クラスタ0	37	肌	購入	化粧水	使用	感じ	乾燥	商品	保	効果	リピート	気	香り	今	リピ	毛穴
クラスタ1	5	肌	購入	効果	化粧水	使用	毛穴	感じ	シミ	イオン導入	商品	気	ニキビ	期待	リピ	リピート
クラスタ2	1	香り	購入	肌	風呂上がり	スプレー	リピート	ミスト	使用	乾燥	感じ	化粧水	リピ	半額	匂い	保
クラスタ3	3	肌	化粧水	香り	購入	使用	感じ	今	気	商品	匂い	乾燥	安心	月桃	グリセリン	好き
クラスタ4	4	購入	肌	化粧水	使用	商品	気	値段	感じ	匂い	満足	保	乾燥	効果	今	使用感

- ・「肌」「購入」「化粧水」などのすべてのクラスタで上位に出現している評価項目がある
- ・「イオン導入」「スプレー」「ミスト」のような特定のクラスタで出現頻度の高い評価項目がある

・本研究ではベクトルの素性として、評価項目の出現割合を用いたため、全てのクラスタで共通する評価項目が上位に現れた抽出された単語に重みづけを行うことでより特徴的な語が上位に現れると考えられる.

・抽出された評価項目と、実際に商品の特徴を表す指標を用いて出力結果と比較し提案手法の評価を行う.

・「香り」や「匂い」, 「リピ」と「リピート」のような同等と考えられる評価項目を統合することでクラスタの特徴的な評価項目が判別しやすくなると考える.

参考文献

謝辞

[1]田村亮介,白石絵里奈,浅沼爽汰,藤田和成,白井聡一,町田翔,延澤志保.類似アプリ比較のためのレビュー自動分類.情報処理学会第81回全国大会(2019)
[2]山西良典,古田周史,福本淳一,西原陽子.出現頻度と構文特徴を用いたレビュー構造の俯瞰のための評価視点の抽出.知能と情報.Vol.27,No.1,pp501-511(2015)
[3]小林 のぞみ, 乾 健太郎, 松本 裕治, 立石 健二, 福島 俊一 “意見抽出のための評価表現の収集”, 自然言語処理Vol.12, No.3, pp.203-222, (2005)
[4]楽天グループ株式会社 (2020): 楽天市場データ. 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.2.1>

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から提供を受けた「楽天データセット」
(https://rit.rakuten.com/data_release/) を利用させていただきました。