

類似アプリ比較のためのレビュー自動分析

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*

*東京都市大学知識工学部

*東京都市大学大学院工学研究科

1 はじめに

Amazon 等のインターネットショップや食べログ等の飲食店レビューサイト、AppStore 等のアプリダウンロードサービス (アプリストア) 等では、ユーザがレビューを投稿したり閲覧したりする機能が提供されているものが多い。本研究では Apple 社の AppStore 内のレビューを対象として、効率的にユーザが利用したいアプリを見つけられるよう、レビューを基に各アプリの特徴を提示し比較検討を支援する手法を提案する。

2 先行研究

稲益は、コスメレビューを対象として、それぞれの評価観点と対応する評価表現の極性を調べ抽出することで、商品の特徴とその評判を自動推定する手法を提案した [1]。相川らは価格.com を対象に商品カテゴリーを用いて比較対象の商品レビューから評価項目を獲得し、それぞれの評価を表にまとめて出力した [2]。平山らは Amazon 商品レビューをモデル化し、商品の評価の基準となる評価属性とそれについての意見や感想となる評価極性を係り受けから抽出し各観点の評価の可視化を行なった [3]。

3 提案手法

スマートフォン向けアプリでは、詳細な機能説明や細かなカテゴリー分類はなく、類似アプリの比較検討が難しい。アプリの比較検討では、アプリの持つ機能や特徴と、それらに対する評価との双方を考慮する必要がある。アプリの機能説明では機能や特徴がアプリ開発者側の視点でわかりやすく説明されているが、ユーザ側から見た特徴が充分記述されているかは疑問である。特に、アプリの問題点等は機能説明には表れないことが多いが、これがユーザにとって重要な情報であることは否めない。そこで本研究では、アプリの機能説明とレビューの双方を組み合わせることで比較検討に必要な情報を抽出する。

3.1 提案手法の流れ

本研究で提案するシステムは、アプリの検索キーワードを入力とし、アプリ同士の比較表を出力する。検索キー

ワードで絞り込まれたアプリ群を対象に、まず機能説明を用いたクラスタリングを行い、その後レビューを利用して各アプリの特徴抽出を行う (図 1)。

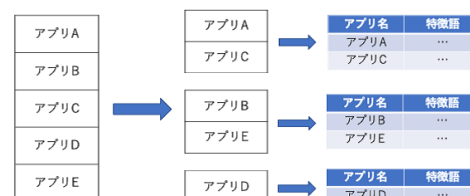


図 1: クラスタリングと特徴語抽出の流れ

3.2 機能説明に基づくアプリのクラスタリング

各アプリについて、AppStore の機能説明に含まれる名詞を抽出し、アプリ群を対象として名詞の TF-IDF 値を用いて x-means 法でアプリのクラスタリングを行う。x-means 法とは初期クラスタ数を設定することで最適なクラスタ数 k を自動で推定しクラスタリングを行う手法で、本研究のように入力によってクラスタ数が変動する場合に有効とされる。

3.3 レビューに基づくアプリの特徴語の抽出

本研究では、アプリの評価を示す特徴語を、対象アプリについてのレビューから抽出する。各アプリに対するレビュー全体から名詞を抽出し、アプリ群に対する TF-IDF 値を基に各アプリの特徴語を上位 5 件まで出力する。

AppStore でのレビュー投稿は、レビュータイトル、レビュー本文、5 段階レーティングを含む (図 2)。レビュー

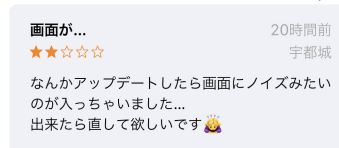


図 2: AppStore のレビュー例

タイトルはレビュー意図に直結することが多いだけでなく、レビュータイトルとレビュー本文が文章として繋がっている場合があるため、本研究では、レビュータイトルとレビュー本文を連結して 1 レビューとして扱う。

3.4 特徴推定における名詞の扱い

クラスタリングや特徴語抽出を行う際の名詞の抽出には、日本語形態素解析システム MeCab で定められている品詞分類を利用した。記号と英数字 (ascii 33-126)、絵文字 (unicode) はノイズとなるため除去した。ノイズ除去のため、特徴語の対象とする名詞は、品詞サブカテゴ

Automatic Review Analysis for the Comparison of Similar Apps
Ryosuke Tamura*, Erina Shiraishi*, Sota Asanuma*, Kazumasa Fujita*, Souichi Shirai*, Sho Machida*, and Shiho H.Nobesawa*

* Faculty of Knowledge Engineering, Tokyo City University

* Graduate School of Engineering, Tokyo City University

りが一般, サ変接続, 形容動詞語幹, 固有名詞, 自立の5種類のいずれかに該当し, かつ, 文字数が2文字以上のもののみとした。

3.5 特徴語の評判推定

レビューで特徴として挙げられた語は, 必ずしも, 肯定的に扱われているとは限らない。そこで本研究では, レビューのレーティングを用いて特徴語の評価を推定する。特徴語のレーティングスコアは, 対象アプリレビューのうちこの特徴語を含むレビューについてのレーティングのみを抜き出し, その平均とする。レーティング平均値は最小値1.00, 最大値5.00であり, これを1点台, 2点台, 3点台, 4点以上の4段階で出力に反映する。

4 実験結果

4.1 実験対象アプリ群

提案手法の有効性を検証するため, 入力キーワードを「カメラ」としてアプリの検索を行い, これによって得られたアプリ193個のうち, レビュー件数が500件を超える49個を比較検討対象アプリ群として, 類似アプリの特徴抽出実験を行なった。ここでレビュー件数でアプリを絞るのは, レビューを基に比較検討を行う手法のため, 十分なレビューが得られることが前提とする。AppStoreのAPIのレビュー取得件数の上限は500件なので, ここではレビュー数500件を絞り込みの閾値とした。それでも比較検討対象アプリ群は49個のアプリを含み, 十分な候補を有していると考えてよい。

4.2 システム出力結果

対象アプリ群に対して特徴語の抽出とレーティングスコア算出を行なった結果を図3に示す。図3では, 各アプリの特徴語をTF-IDF値に基づき上位5件出力している。それぞれの特徴語はレーティングスコアによって4段階で色分けされており, 各特徴語が評価の極性(好意的な評価か否か)も含めて直観的に検討できる。また, クラスタごとに出力するため, クラスタ全体の特徴も読み取ることが可能である。そのため, アプリ群全体をひとつひとつ比較検討せずとも, まずクラスタ単位で大きく見比べた後, クラスタ内で類似アプリを比較する, という2段階でアプリ選択が可能である。

本研究では, アプリ群のクラスタリングで用いるx-means法の初期値クラスタ数を15として実験を行った。その結果, 49個のアプリ群が16種類のクラスタに分類されたが, 複数のアプリを含むクラスタは3個のみで, 他は1アプリ1クラスタとなった(図3)。1番大きなクラスタは18個のアプリ, 2番目のクラスタは15個, 3番目のクラスタは3個から成る。1番クラスタは便利さを

クラス	App Name	App ID	平均値	特徴語1	特徴語2	特徴語3	特徴語4	特徴語5
1	ミール	472973118	4.77	美味	料理	寿司	夜飯	ミール
1	セルフタイマーカメラ	526262545	4.69	タイマー	セルフタイマー	セルフ	シンプル	機能
1	簡便カメラSnapfish	436711212	4.68	機能	シンプル	シンプル	機能	操作性
1	Epica 2	911219705	4.37	カメラ	おもしろ	面白	おもしろ	つよし
1	Epica	788385091	4.27	面白	面白	おもしろ	おもしろ	リアル
1	超簡単カメラ	446278045	4.15	簡単	簡単	シャッター	シャッター	簡単
1	OneCam	422845617	4.01	簡単	カメラアプリ	シャッター	シャッター	操作性
1	簡単カメラ	569723994	3.80	簡単	カメラ	簡単	簡単	操作性
1	Photograph+	1028791881	3.77	簡単	簡単	簡単	簡単	操作性
1	Simple Resize	327776379	3.64	リサイズ	簡単	簡単	簡単	操作性
1	COOLLOLO	973325431	3.56	簡単	簡単	簡単	簡単	操作性
1	Frameio	568780324	3.53	フレーム	簡単	簡単	簡単	操作性
1	StateCameraHD	912177018	3.39	簡単	簡単	簡単	簡単	操作性
1	PICSLAY 2	942044355	3.27	簡単	簡単	簡単	簡単	操作性
1	ストレススキャン	673503525	3.15	簡単	簡単	簡単	簡単	操作性
1	カメラ360	443354861	2.71	簡単	簡単	簡単	簡単	操作性
1	TiTok	1235601864	2.44	簡単	簡単	簡単	簡単	操作性
1	SNOW	1022267439	2.42	簡単	簡単	簡単	簡単	操作性
2	PrimeCamera	1225971987	4.46	簡単	簡単	簡単	簡単	操作性
2	ゆるふわ美肌加工	1028887806	4.73	簡単	簡単	簡単	簡単	操作性
2	8ミリカメラ	409541444	4.68	簡単	簡単	簡単	簡単	操作性
2	壁紙サイズに加工	863094731	4.46	簡単	簡単	簡単	簡単	操作性
2	isoCamera	980538523	4.32	簡単	簡単	簡単	簡単	操作性
2	SweetCamera	1241681287	4.27	簡単	簡単	簡単	簡単	操作性
2	YouCam	963844475	4.05	簡単	簡単	簡単	簡単	操作性
2	Microsoft Pix カメラ	1127910488	3.82	簡単	簡単	簡単	簡単	操作性
2	簡便カメラ	1394426952	3.56	簡単	簡単	簡単	簡単	操作性
2	Piozoo	956441499	3.46	簡単	簡単	簡単	簡単	操作性
2	女神カメラ	539967602	3.46	簡単	簡単	簡単	簡単	操作性
2	B612	904209370	3.10	簡単	簡単	簡単	簡単	操作性
2	Foodie	1076859004	3.05	簡単	簡単	簡単	簡単	操作性
2	BeautyPlus	622434129	1.78	簡単	簡単	簡単	簡単	操作性
2	Instagram	389801252	1.68	簡単	簡単	簡単	簡単	操作性
3	簡便カメラ	357883632	3.62	簡単	簡単	簡単	簡単	操作性
3	簡単カメラ	578508665	4.24	簡単	簡単	簡単	簡単	操作性
3	簡便カメラ	919409463	3.81	簡単	簡単	簡単	簡単	操作性
4	シンプルカメラ高画質	1048899002	3.36	簡単	簡単	簡単	簡単	操作性
5	Google 翻訳	414708508	1.84	簡単	簡単	簡単	簡単	操作性
6	どこでもカメラ	989898913	4.39	簡単	簡単	簡単	簡単	操作性
7	アメリカメラ	784201986	3.90	簡単	簡単	簡単	簡単	操作性
8	簡便カメラ	1079316689	2.86	簡単	簡単	簡単	簡単	操作性
9	有名カメラ	1122366064	2.44	簡単	簡単	簡単	簡単	操作性
10	簡便カメラ	487074164	2.11	簡単	簡単	簡単	簡単	操作性
11	スマートフォン	845133175	1.84	簡単	簡単	簡単	簡単	操作性
12	簡便カメラ for iPhone	647789839	3.24	簡単	簡単	簡単	簡単	操作性
13	My Heart Camera	657224393	4.27	簡単	簡単	簡単	簡単	操作性
14	超簡単 for iPhone	1174021823	3.93	簡単	簡単	簡単	簡単	操作性
15	Pix カメラ	1062189449	3.11	簡単	簡単	簡単	簡単	操作性
16	簡便カメラ	768917254	4.43	簡単	簡単	簡単	簡単	操作性

図3: システム出力結果

強調しているアプリ, 2番クラスタは加工を強調しているアプリ, 3番クラスタは特殊加工を強調しているアプリが多いが, 十分綺麗に分かれているとはいえ, クラスタリング精度の向上については検討の余地がある。

また, 特徴語のレーティングスコアについては, 特に有名かつ人気の高いアプリで明らかに値が低いという特徴が見られた。例えば, カメラアプリとして有名なSNOWやBeautyPlus等は, 人気アプリであるにも関わらずレーティングスコアがかなり低い。これは, 人気の低いアプリではそのアプリを応援するようなレビューが多くレーティングスコアが高くなる傾向があるのに対して, 既に人気の高いアプリでは問題点の指摘や期待外れといったレビューが増えることに起因すると考えられ, レーティングスコアのみで頼らない適切な極性評価方法についても検討すべきと考えられる。

5 まとめ

スマートフォン向けアプリストアではアプリの検索や絞り込み機能が少なく, 検索時に候補となる大量のアプリの中からユーザーが欲しいと思う機能や特徴を持つアプリを素早く見つけるのは難しい。本研究では機能説明とレビュー内容からアプリの特徴を明確にした上で人気度の高いアプリを推定する手法を提案した。

参考文献

- [1] 稲益 ほなみ, “評価視点に着目した口コミサイトの評価の視覚化,” 東京都市大学卒業論文, 2012.
- [2] 相川 直視, 山名 早人, “レビューからの商品比較表の自動生成,” 言語処理学会 第17回年次大会発表論文集, pp.512-515, 2011.
- [3] 平山 拓央, 湯本 高行, 新居 学, 佐藤 邦弘, “語の共起と極性に基づく商品レビュー閲覧支援システム,” 情処研報 Vol.2012-DBS-155 No.3, pp.1-9, 2012.