

研究指導 中澤 真 教授

# スマートフォンアプリケーションの特徴と評価の的確な抽出手法

—アプリストア内レビューに含まれる品詞の係り受け関係に着目して—

木村 佳月

## 1. はじめに

スマートフォンの世帯別保有割合は、2015年時点で72.0%だったのに対し、2020年時点で86.8%と増加傾向が続いている[1]。スマートフォンの利用者数の伸びに合わせて、スマートフォン向けに提供されるアプリケーション数の増加も顕著になっている。App Annieが2022年に公開した「モバイル市場年鑑2022」では、2015年のアプリケーションの累計リリース数は720万本だったものが、2020年には1,935万本となっている[2]。アプリケーション数の増加は、類似の機能を持つアプリケーションの増加にもつながり、ユーザの選択肢の幅を広げる一方で、自分に適したアプリケーションを選ぶのを難しくしているという側面も持っている。

大量のアプリケーションから、ユーザがアプリケーションを選択する際、参考となるのはアプリケーションのダウンロード数やレビュー数、レビュースコアなどである。しかし、それらからはアプリケーションの優劣を把握することはできるものの、類似している中でどのような違いがあるのかは掴むことができない。それぞれのアプリケーションの特徴を理解するには、レビュー内容を詳細に読む必要があるが、大量のデータの中から必要な情報を探し出すのはかなりの労力を必要とする。

この問題に対するユーザ支援のために、特徴的な情報を自動抽出するための研究が取り組まれている[3][4][5]。アプリケーションの特徴は機能を表す語句と、それに対する評価を表す語句で構成されているため、機能と要素をそれぞれ抽出することが必要になる。しかし、現状としては機能を表す語の抽出はできているものの、それに対応する評価を表す語の抽出はできていない。

これを解決すべく本研究では、レビュー内に存在する品詞の係り受け関係を用いて、アプリケーションの特徴とこれに対応する評価を的確に抽出する手法を提案する。

## 2. アプリケーションとアプリストア

### 2.1 スマートフォン向けアプリケーションの現状

2020年6月現在、日本国内におけるスマートフォン

の世帯別保有割合は86.8%[1]となっており、スマートフォンの利用者は増加の一途を辿っている。またスマートフォンは、年々スペックや機能の向上がなされている。

それに伴い、アプリケーションの数も年々増加傾向にある。スマートフォン向けOSはAppleの提供するiOSとGoogleの提供するAndroidに大別され、それぞれのアプリストア(App Store, Google Playストア)において提供されるアプリケーションの総数は、2021年時点で合計540万本程<sup>1</sup>となっている[2]。アプリケーションは国境を越えて配信されるため、世界中にアプリケーションを展開したいという開発者や企業がそれらを開発・配信している。その結果、2021年のアプリケーションの新規インストール数は2,300億回を優に超えている[2]。これらのことから、スマートフォン向けのアプリケーションの需要は非常に高いといえる。

### 2.2 アプリストアの現状

膨大な種類のアプリケーションを提供するアプリストアには、キーワード検索機能や、カテゴリ検索機能があるが、このような機能を用いても類似したアプリケーションが多数表示されてしまう現状にある。例を挙げると、Google Playストアにて「音楽プレーヤー」という語で検索を行うと、その言葉に該当するアプリケーションが243件<sup>2</sup>表示される。この一覧の表示順は、アプリケーションのダウンロード数やレビュー数、レビュースコアなどで決定される[6]が、最初の方に提示されるアプリケーションが必ずしも自分に適したアプリケーションであるとは限らない。上位表示されるのはあくまでも多数のユーザから評価を得ているアプリケーションであり、評価が少ないアプリケーションにユーザの求める機能が含まれている可能性もある。このため、ユーザはこれらの情報とは別にレビューの内容を読むことで、その中から自分が求めるアプリケーションの特徴を探さなければならない。

一般的に特徴を表す情報は、アプリケーションに特有の機能を示す機能要素と、それを評価する評価表現で構成される。本研究では、機能要素をアプリストア内のレビューに含まれるアプリケーションの機能を表す語または語句、評価表現は機能要素を説明・表現する語または語句と定義する。例として、「写りは

<sup>1</sup> 実際に配信されているアプリケーション数は、リリース数より少ない。これは配信を終了したアプリケーションが存在するためである。

<sup>2</sup> 2022年2月時点

綺麗なので、良いアプリだと思います。」といったレビューが存在した場合、機能要素は「写りは」となり、評価表現は「綺麗で」となる。これは、「写り」というアプリケーションの特徴を、「綺麗」と評価しているためである。一方、例内の語「アプリ」は単独では機能の意味を持たず、「良い」については対応する機能要素が存在しないため、いずれも機能要素や評価表現にはならない。

しかし、アプリケーションごとに寄せられているレビューを1件ずつ確認する労力はかなりのものになってしまうため、それぞれの機能要素と評価表現をレビュー文からの的確に提示できるシステムが求められている。以上のことから本研究では、アプリストア内に存在するアプリケーションに寄せられているレビューから、アプリケーションの有する機能要素と、それと対となる評価表現を抽出する手法を探る。

### 3. アプリケーションの特徴抽出手法

#### 3.1 先行研究と問題点

投稿されているレビューからアプリケーションの特徴を抽出する研究として、高増ら[3]はアプリストア上で検索を行った結果から共起ネットワーク[7]を構築、その中から重視したい特徴を選択させ、それと共起する語を機能要素として抽出する手法を用いている。田村ら[4]では、アプリケーションごとにTF-IDF[8]値が高い語を抽出し、そのレビューの5段階スコアが高い場合にはポジティブな特徴を表す語として、逆にスコアが低い場合にはネガティブな特徴を表す語としてユーザに提示する手法を提案している。

しかし、これらの先行研究には問題がある。高増らでは、機能要素に対する評価をポジティブかネガティブかという二極化した指標で示している。その関係上、「少し良い」というものや「かなり良い」というものでも同じポジティブに振り分けてしまう。これではアプリケーションの機能要素が、どの程度の評価をレビュー内で得ているのか、ユーザや開発者が把握しにくくなってしまふ。一方の田村らの手法は、機能要素に対する評価としてレビューに付与されたレビュースコアを用いているため、「Aの機能は良いがBの機能はダメだ。」のような複数の機能要素に関するレビューに対して、適切な評価ができなくなってしまうという問題がある。

そこで、本研究では特徴量としてレビュー内の係り受け関係と品詞、語の出現頻度に着目する。係り受け関係を見ることで、同一レビュー内に複数の機能要素が存在する状態であっても、それに対応する適切な評価表現を取り出すことが期待できる。その上、評価表現を取り出すことで機能要素がどの程度の評

価を受けているのかをユーザが確認することができる。また、語の出現頻度に着目することで、アプリケーションの特徴を表す情報を高精度に抽出できる可能性がある。これらの特徴量とすることで、レビュー内の機能要素と評価表現の的確な抽出を試みる。

#### 3.2 本研究で行う分析の手法

本研究では、分析に用いるアプリケーションの全レビューから、機能要素を含む文節と評価表現を含む文節を抽出する手法を提案する。

なおデータとして、アプリストア内のアプリケーションに寄せられているレビューを用いる。レビューの収集にはWebスクレイピングソフトであるOctoparse 8<sup>3</sup>を用いる。収集条件は、2021年11月22日にGoogle Playストアで「カメラ」というキーワードで出現するアプリケーションとし、なおかつレビュー件数上位の中から選定した1つに含まれるレビュー200件を本研究で用いる。

まず、収集したレビューから機能要素と評価表現の対を人手により261件取り出した。これらがシステムで抽出を目指す正例データとなる。この正例を形態素解析器のMeCab<sup>4</sup>を用いて分析し[8]、機能要素と評価表現の品詞傾向を明らかにする。表1は該当する品詞を含む文節数を示したものである。

表1 機能要素と評価表現の品詞別の出現文節数比較

(単位:文節数)	機能要素	評価表現
名詞	255	100
動詞	48	221
形容詞	10	66
形容動詞	20	21
助詞	231	98
助動詞	12	144

分析の結果、機能要素の正例に最も多く含まれる品詞は名詞、評価表現に最も多く含まれる品詞は動詞となったため、これらを本研究で取り扱う品詞とする。なお、評価表現に関しては形容詞も対象とする。理由として、形容詞は「綺麗」や「ダメ」などの、物事の状態・評価を表すことが多いためである。

機能要素は名詞を含む文節で評価表現は形容詞もしくは動詞を含む文節という条件のみでは対象があまりにも多くなるため、適合率[8]が極めて低くなってしまふ可能性がある。そこで、正例における機能要素と評価表現の係り受け関係や距離の傾向を明らかにして、抽出範囲を限定できるようにする。なお、距離とは機能要素と評価表現がどれほど離れているかを示す指標であり、隣接する場合は1となる。評価表現に係り受け関係がある対の件数、対となる語の距離が2以下の件数、距離が3の場合の件数を分析し

<sup>3</sup> <https://www.octoparse.jp/>

<sup>4</sup> <https://taku910.github.io/mecab/>

た結果を表 2に示す. なお, 係り受けの解析には CaboCha<sup>5</sup>を用いた.

表 2 正例内の関係性別の該当件数

		件数
正例の総数		261
係り受け関係		227
距離関係	~2	233
	3	15

この結果から, 機能要素と評価表現の対の多くは係り受け関係があり, 出現位置も距離が2以内のごく近い範囲であることが判明した. そのため本研究では, 機能要素として名詞を含む文節を, 評価表現として動詞または名詞を含む文節を一对とし, 係り受け関係と距離を用いて検証する.

#### 4. 抽出手法の検証結果

前項で求めた正例に含まれる品詞傾向より, 機能要素には名詞が, 評価表現には形容詞や動詞が多く含まれることが判明した. また, それらの関係性として係り受けがあり, かつ文節同士が近い距離にあるものが多いことも判明した. そこで, これらの条件をどのように組み合わせることが, 抽出精度の向上に寄与するのか明らかにすることを考える. 組み合わせとしては以下の5つの仮説を用意し, それぞれの条件を満たすものを機能要素と評価表現の対であると判別したときの精度を分析する.

- 仮説1. 係り受け関係にある名詞と形容詞の対
- 仮説2. 係り受け関係にある名詞と動詞の対
- 仮説3. 仮説1と仮説2の両方を組み合わせたもの
- 仮説4. 仮説3の条件に加えて, 係り受け関係にある対の文節同士が隣接するもの
- 仮説5. 仮説3の条件に加えて, 係り受け関係にある対の文節同士が2文節以内に存在するもの

表 3 仮説別の正解率と適合率と再現率

	正解率	適合率	再現率
仮説1	16.88%	34.42%	20.31%
仮説2	19.87%	37.95%	24.14%
<b>仮説3</b>	<b>51.99%</b>	<b>52.11%</b>	<b>90.04%</b>
仮説4	43.37%	43.96%	73.95%
仮説5	50.22%	50.44%	86.97%

表 3に示した検証結果より, 仮説1と2の結果はすべての値において低いものだったが, この2つを組み合わせた仮説3では再現率が約90%と高い精度となった. 一方で仮説3の条件に文節の距離を追加した, 仮説4および仮説5では, 仮説3と比較してすべての値が減少する結果となったことから, 文節の距離は機能要素と評価表現の抽出に適さないと結論付けた.

仮説3を用いた場合, 正例をどれだけ網羅的に抽出できているかという再現率は約90%と高いが, 適合率の結果からは機能要素と評価表現の対と判別したものの約半数が誤っていることになっており改善の必

要がある. そのためには, ノイズとなっている不正解データと, 正解データの差異を表す適切な特徴量について検討する必要がある.

#### 5. 新特徴量の構築と検証

##### 5.1 差異係数の導出

適合率を高める特徴量を把握するために, 仮説3において機能要素と評価表現の対であると判別したデータを, 正解データ235件と不正解データ216件の2群に分けることから始める. これら2群のデータに含まれる語と品詞を, テキストデータ解析ソフトウェアの KH Coder<sup>6</sup>を用いて, 正解データおよび不正解データに含まれる語および品詞の出現頻度と差異係数 [9] を分析する.

$$\text{差異係数} = \frac{\text{データAの頻度} - \text{データBの頻度}}{\text{データAの頻度} + \text{データBの頻度}}$$

差異係数は-1以上1以下の範囲の値をとり, データAに偏って出現する語の場合は1に近い値となり, 反対にデータBに偏って出現する語の場合は-1に近い値となり, 両群に同じように出現している語の場合は0に近い値となる. このため, 差異係数の絶対値の大きな語や品詞ほど, 2つの群を判別するのに有用なものと考えられることができる.

品詞ごとに出現頻度と差異係数を求めた結果は表 4に, 語ごとに出現頻度と差異係数を求めた結果は表 5に示す.

表 4 品詞別の出現頻度数と差異係数

品詞	正解データ 出現頻度	不正解データ 出現頻度	差異係数
形容動詞	101	49	0.35
副詞可能	47	23	0.34
否定助動詞	102	51	0.33
人名	2	1	0.33
名詞	392	234	0.25
形容詞	67	41	0.24
感動詞	3	2	0.20
サ変名詞	315	212	0.20
副詞	7	5	0.17
動詞	192	147	0.13
形容詞 (非自立)	22	19	0.07
ナイ形容	4	9	-0.38
固有名詞	1	3	-0.50

表 5 抽出語別の出現頻度数と差異係数 (上位 10 語抜粋)

差異係数が1に近いもの				差異係数が-1に近いもの			
抽出語	正解	不正解	差異係数	抽出語	正解	不正解	差異係数
ピント	15	0	1	タップ	0	5	-1
合う	11	0	1	XPERIA	0	4	-1
項目	6	0	1	バッグ	0	4	-1
使い物	7	0	1	つける	0	3	-1
再生	3	0	1	オフ	0	3	-1
指定	5	0	1	タイミング	0	3	-1
翻訳	3	0	1	書く	0	3	-1
とれる	3	0	1	無効	0	3	-1
レート	3	0	1	無い	1	6	-0.71429
英語	4	0	1	お願い	1	5	-0.66667

<sup>5</sup> <https://taku910.github.io/cabochoa/>

<sup>6</sup> <https://kncoder.net/>

両群での出現頻度が100以上かつ差異係数のある程度高い品詞が判別に有効であると考え、表 4の結果から否定助動詞と形容動詞の有無を特徴量として用いることにする。

一方、抽出語については表 5に示したように出現頻度が少ないため、この時点では出現頻度と差異係数の適切な閾値を定めることができない。それゆえ、差異係数については0.3と0.2、出現頻度については3, 4, 5の閾値を変化させて、もっとも精度の高いケースを明らかにする。ただし、差異係数が正となる語と負となる語の両方が存在するため、特徴量としてはこれを分けて考える。なお、差異係数が正の値で用いることになった語を正解特徴語、逆に負の値となったものを不正解特徴語と今後呼ぶことにする。

### 5.2 数量化 II 類を用いた検証

前項までで決定した特徴量を用いて、判別の中率が最も高くなる組み合わせを明らかにするために、数量化II類[10]を用いて分析する。その際、利用する特徴量は以下の4種であり、対象文節内での出現の有無となる。これらの特徴量を4つの説明変数とし、目的変数は対象データが機能要素と評価表現の対となっているか否かを示すラベルとする。

- 否定助動詞の有無
- 形容動詞の有無
- 正解特徴語の有無
- 不正解特徴語の有無

差異係数および語の出現頻度別の判別の中率を表 6に示す。

表 6 差異係数の閾値と出現頻度の閾値の組み合わせによる判別の中率

差異係数閾値	出現頻度閾値	判別の中率
0.3	5	73.17%
	4	73.84%
	3	<b>75.17%</b>
0.2	5	70.51%
	4	71.40%
	3	73.84%

分析結果より、差異係数の閾値が0.3、語の出現頻度閾値が3の組み合わせの場合に、最も高い判別の中率となることが明らかになった。また、出力形式としては以下の表 7に示すものとなり、先行研究と比較して詳細な評価の抽出ができた。

表 7 先行研究と比較した本研究の出力結果の違い

	機能要素	評価表現
先行研究	セルフタイマー	ポジティブ
本研究	セルフタイマーが	長くていい。

結論として、アプリレビュー内の機能要素と評価表現の抽出には、名詞を含む文節と、それと係り受け関係になっている形容詞または動詞を含む文節の

対を抽出し、特定の品詞や正解特徴語を含まない文節、不正解特徴語含む文節を除外する手法が有効だといえる。

## 6. むすび

先行研究では、抽出した特徴に対して係り受け語やレビュースコアからポジティブかネガティブの推定を行っていたが、評価表現と共に抽出する手法は用いていなかった。そこで本研究では、係り受け関係にある名詞と形容詞または動詞の対と、4種の特徴量から機能要素と評価表現を抽出する手法を考案した。その結果、先行研究でのポジティブやネガティブなどの二極化した評価ではなく、より詳細な評価表現を一定の精度で抽出することに成功した。これはアプリストア上でアプリケーションの特徴をシステムが自動提示するための第一歩となる。また、一般ユーザがアプリケーションを探しやすくなるだけでなく、開発側にとってもアプリケーションに対する評価を効率的に把握するのにも役立つはずである。

しかし、本研究においては単一のアプリケーションに対してのみ検証を行っていたため、今後はアプリケーションの数と種類を増やしての抽出精度向上や、異なるアプリケーションのレビュー文を比較することによる特徴抽出について検討を続ける必要がある。

## 参考文献

- [1] 総務省, 令和 2 年通信利用動向調査, <http://urx3.nu/TUaZ> (参照 2022-02-07)
- [2] App Annie, モバイル市場年鑑 2022: 消費者が 3.8 兆時間をモバイルデバイスに費やすモバイルファーストな新時代における市場攻略法, <http://urx3.nu/VqEp> (参照 2022-02-07)
- [3] 高増 広大, 田原 康之, 清 雄一, 大須賀 明彦, “ユーザのこだわりを考慮したスマートフォンアプリケーション推薦システムの提案”, 研究報告ソフトウェア工学(SE) 2016-SE-191 No.2, pp.1-8, 2016.
- [4] 田村 亮介ほか, “類似アプリ比較のためのレビュー自動分析”, 第 81 回全国大会講演論文集, Vol.2019, No.1, pp.125-126, 2019.
- [5] 安倍 小百合, 小林 一郎, “レビューデータにおける推薦のための属性抽出”, 第 78 回全国大会公演論文集, Vol.2019, No.1, pp.513-514, 2016.
- [6] Apple, 検索に向けた最適化, <https://developer.apple.com/jp/app-store/search/> (参照 2022-02-07)
- [7] 末吉 美喜, テキストマイニング入門 Excel と KH Coder でわかるデータ分析, オーム社, 2019.
- [8] 金 明哲, テキストアナリティクスの基礎と実践, 岩波書店, 2021.
- [9] 小林 雄一郎, ことばのデータサイエンス, 朝倉書店, 2019.
- [10] 石井 俊全, 意味がわかる多変量解析, ベレ出版, 2014.