

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

高増広大^{†1} 田原康之^{†2} 清雄一^{†3} 大須賀昭彦^{†4}

概要：近年、スマートフォンの普及率が世界的に増加しているのに付随してスマートフォンアプリケーションの需要が高まっており、開発が進められてきている。そのため同じ課題を解決できるアプリケーション、すなわち類似したアプリケーションが多数出現している。そこでユーザはアプリケーションを利用して目的を満たす事は最低限の条件として、アプリケーションのコンテンツや使い易さ等、自分がこだわる観点でのアプリケーションの選定を行いたくなると考えられるが、多数の中からの選定は困難である。本研究では、ユーザが目的を達成する為にアプリケーションを導入するという前提で、ユーザのこだわりを考慮した推薦システムを提案する。そのために、通常アプリケーションの入手先となるアプリストアに投稿された他のユーザのレビューコメントに自然言語処理手法を適用する。具体的には、コメントから類似アプリケーションの特徴を表すと考えられる単語を抽出し、ユーザはそこから興味を持った単語を選ぶ。アプリケーションのレビューコメントにおいて単語の評価がネガティブかポジティブかを判定することにより、アプリケーションのその単語に関する評価値を算出し、ランク付けして推薦する。算出した評価値を、実際にアプリを使用した被験者による評価と比較し、またレビュー内に出現する単語に対するネガティブ・ポジティブの判定を、人手で作成した正解データと照らし合わせる事でシステムの評価を行った。

キーワード：スマートフォン、アプリケーション、ユーザ、レビュー、情報推薦、テキストマイニング

1. はじめに

近年、スマートフォンの普及率が世界的に増加しているのに付随してスマートフォンアプリケーションの需要が高まり、開発が進められてきている。スマートフォンアプリケーションは、ユーザの日常生活における課題の解決から、娯楽の提供、ソーシャルネットワークサービスによる交友関係の管理、勉強、と幅広い場面で活躍しており、もはや必要不可欠となっていくのではないかと考えられている。その結果、アプリマーケットのダウンロード数は累計100億ダウンロードを超えているという現状から、市場が広がっていると共にユーザはアプリケーションに対する関心は日に日に高まっていると言える。アプリストアに配信されている数も、iOS向けのアプリケーションは50万以上、Android向けのアプリケーションは25万以上と、膨大な数のアプリケーションが登録されていると報告されている。また、近年ではアプリケーションのビジネスが注目されると共に、国境を超えて自由にアプリケーションの開発及び配信をする事が可能であり、これからはますますアプリケーションの配信数及び、開発者が増加していくことが予想される。

アプリケーションは基本的にAppStore、Google Play等に代表されるアプリストアからダウンロード及びインストールを行う事でスマートフォンに導入する形態を取る。アプリストア上のコミュニティにアプリケーションに対するユーザの評価を記載し、ユーザはアプリケーションの課題や更なる改善の展望を開発者及び、他のユーザに提供する事でアプリケーションの更なる発展及び、ユーザに恩恵を与

えていく事に繋がっているのが現状である。

一方で、アプリケーションの登録数が増えると共に、アプリケーションの機能が類似したものが出現していくと考えられる。アプリストア上でキーワードを入力して、アプリケーションの検索を行った際、膨大な数の結果が表示される事も珍しくない。どのアプリケーションをダウンロードするかを決める上で判断材料となる情報は、アプリストア上に表示される各アプリケーションの説明文やダウンロード数、平均評価値、投稿レビューといくつかあげられるが、情報量が多く、参考にして良いか分からないユーザも少なくないのが現状である。また、アプリケーションの利用する決め手の判断材料となる、ダウンロード数の多さ、既に利用しているユーザによるレビューの平均評価が高い等の判断材料は必ずしも有効でない場合もある。ユーザは今後、目的を満たすのは最低限の条件として、アプリケーションのコンテンツや使い易さ、デザイン等が自分に向いているかどうか等の質にこだわったアプリケーションの選定が必要として求められていると考えられる。

こだわりを抽出、分類した上で評価及び比較を行う研究は未だ存在していない。本研究では、ユーザがアプリケーションを導入する上で、ユーザに合わせた質を考慮した推薦システムを提案する。本システムは、ユーザがキーワード検索などで一群の類似アプリを特定すると、自然言語解析ツール KH Coder[1]を利用して、アプリストアに投稿されたレビューから、類似アプリに特徴的な単語を抽出し、共起ネットワークとして視覚的にユーザに提示する。そしてユーザはその単語群の中から、自分のこだわりに関連する単語を選ぶと、各アプリのレビュー内において、それらの単語に対して表現されているネガティブ・ポジティブの判定を行う。そのために、あらかじめネガティブ・ポジティブ

^{†1}~^{†4} 電気通信大学
The University of Electro-Communications

プのラベル付けを人手で行った辞書を用意する。そして各レビューに対し係り受け解析器 CaboCha[2] を適用し、ラベル付けされた単語にユーザが選んだ単語が係っているかどうかを調べることにより判定を行う。その上で判定に応じてレビューに付けられた5段階評価値を再評価した値を、ユーザが選んだ単語に対するそのレビューの評価値とする。最後に全レビューの評価値の平均により各アプリの評価値を算出し、ランク付けしてユーザに提示する。このように、アプリのレビューに対し自然言語処理手法を適用することにより、ユーザのこだわりを考慮したアプリ推薦を行う点が、本研究の貢献である。

本論文の構成は次の通りである。第2章では関連研究を紹介する。第3章では提案手法を説明する。第4章では実験内容について、第5章では結果について記述する。第6章で考察を述べ、第7章でまとめと今後の課題を示す。

2. 関連研究

本研究はアプリケーションの投稿レビューを基にユーザへのアプリケーションの推薦を提案する。先行研究として、投稿レビューを基に、推薦を行う関連研究は既に盛んに行われており[3]、ユーザが投稿するレビュー文書に注目し、投稿されたレビューから、ユーザの観点と評価の関係を要約している研究[4]などが存在する。レビューから評価観点に着目した推薦を行うにあたって、推薦する物の価値や、ユーザの評価観点の基準に依存して、推薦の有効性が変わると報告されていることから、推薦する上で、多様な視点での提案方法が求められていると考えられる。そもそも、レビューが意味する事柄としては、商品やシステムを実際に利用したユーザの感想が含まれているのが周知の事実だが、一言に感想といえども、使用感想から、レビューの読者へのメッセージ、今後の期待など、様々な情報が含まれることが分かっており、安藤[5]らはレビューの内容と、レビューを確認した人への購買意欲に関わる影響の度合いに、関連性がある事を報告している。ソフトウェアに関わる、投稿レビューはユーザによる需要と開発者の供給の構図から、今後の改善に関して非常に重要な情報である。A. Begel[6]らは、マイクロソフトの社員はソフトウェアの開発を行うにあたって、ユーザは製品・システムを利用するにあたって、どの部分を評価してくれているか?という疑問を大半が抱いているという報告がされている。Ning[7]らは投稿レビューを基に、開発者へのアプリケーション開発における課題提示を行う事で、開発改善に貢献した。研究の成果及び精度としては、有効性がある結果を示した一方で、開発者への提示という知見の提供先を考慮すると、既に把握している情報も当然あり、知見としては重要な情報を得る事の難しさを示した。また、アプリケーションの推薦ではないものの、小川[8]らはユーザの購買情報を用いた、

推薦アルゴリズムを提案しており、従来の協調フィルタリングを用いた、似たり寄ったりの推薦ではなく意外性を残しつつ、適合性の高い推薦を提案するのに成功している一方で、ユーザが既に知っているアイテムを推薦される事に慣れてしまうと、興味を持たなくなってしまうという事態も同時に抱える事を今後の課題としてあげている。花岡ら[9]は、同じ製品カテゴリである他の商品名を投稿レビューに載せる事で比較及び評価を行うユーザの存在から、商品間を対比する文章を抽出する事で、評判分析によるユーザへの推薦に向けて貢献している。研究の成果としては有効性を強く示すことが出来ず、問題点として、商品のカテゴリの不整合や、ユーザが他の商品を指す際の表現の仕方による精度の変化をあげていた。本研究では、利用目的が同一である複数のアプリケーションの投稿レビューを用いる事で、評価における任意の項目の表現について、記述の揺れ方は少ないと考えた。

3. 提案手法

本章では、本研究におけるユーザへのアプリケーション推薦の提案手法を述べていく。

本研究はどのアプリケーションを利用しても特定の目的を満たす事が出来るアプリ群からユーザに合わせたアプリケーションを推薦する事を目的とする。提案手法の概要を図1に示すと共に、工程を(1)~(5)に記載する。

(1)-ユーザは、目的を達成する為にアプリケーションを導入に向けて、アプリストアで検索を行う。例として、交通機関の時刻表のアプリケーションを導入したい場合は、アプリストアにて「時刻表」というキーワードでアプリケーションの検索を行う。

(2)-検索結果として表示される時刻表に関わる、複数のアプリケーションを対象に、KH Coderによる共起ネットワークが生成され、ユーザに提示される。また、その際、提示された共起ネットワーク上から、アプリケーションを利用する上で重視したい単語を選んでもらう。

(3)-選ばれた単語を基に、評価項目を生成。

(4)-アプリケーションの群の内、各アプリケーションの投稿レビュー内に含む評価項目に属する単語に対して特定のネガティブ・ポジティブの単語の係り受けが発生しているかどうかを分析し、各アプリケーションの評価項目に対する評価値を導出する。

(5)-導出された評価値が高いアプリケーションをユーザへ推薦する。提案手法の詳細として、KH Coderによる共起ネ

ネットワークの生成は 3.1 に、評価項目の生成は 3.2、評価項目に対する評価は 3.3 に示す。

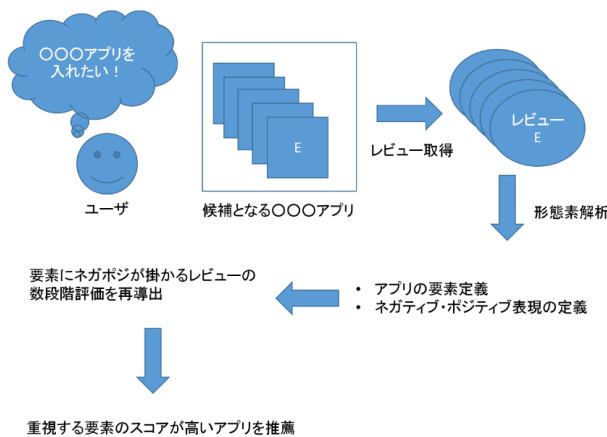


図 1.提案手法全体図

評価項目(以降、機能要素)は、投稿レビューに含まれる、アプリケーションの評価に影響を及ぼす単語として定義する。投稿レビューに、件名及び本文にテキスト情報が記載されている場合、数段階の評価の要因は、テキスト情報に含んでいると想定し、評価に強く影響を与えていると考えられる。

以上から、アプリケーションのユーザによる数段階評価の決定要因は投稿レビュー内の単語が強く関係し、特に頻度が高く出現する単語が評価の決め手になる可能性が高いと考えた。本研究では、投稿レビューの内、本文に出現する単語情報を基に行った。本文中に含む単語の共起情報を基に、機能要素を構成する単語に対して、投稿レビューの文書上から、ネガティブ・ポジティブな意味を持つ単語が係っているかの情報から、元のレビューの評価値とは別に、本研究で考案する評価値を導出する。以上を通して、アプリケーションの評価を構成するいくつかの機能要素の数段階評価をユーザに提示する事でアプリケーションの推薦に繋げるのが研究の流れである。

3.1 KH Coder による共起ネットワークの生成

本研究では KH Coder という Perl で作成されている自然言語処理におけるテキストマイニングのソフトウェアを用いた。KH Coder が持つ機能の中でも、本研究では特定のキーワードを含む文書の抽出、文書内に共起して出現する単語抽出、また共起ネットワークの生成に関して利用した。

3.2 機能要素の生成

機能要素の生成は、ユーザが共起ネットワーク上から選んだ単語が属するサブグラフに依存し、決定される。サブグラフとは、共起ネットワークにおける、グラフ内部で区別

可能な形で相互に結びついた下位集団のノード群であり、色分けで表現される。図 3 は、あるアプリケーション群の投稿レビューから、生成された共起ネットワーク上のサブグラフである。ここでは、単語群{月、グラフ、表示、画面、カレンダー}を機能要素とする。

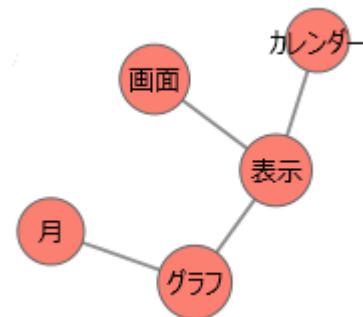


図 2.サブグラフの例

3.3 機能要素に対する評価

機能要素に対する評価は、推薦の候補である各アプリケーションの投稿レビューを用いて導出した。各アプリケーションの投稿レビューの本文全体を形態素解析エンジン MeCab[10]の辞書を用いて、品詞分解を行い、名詞以外の品詞の頻出単語上位に該当する単語を抽出。「顔文字」や、インターネット上におけるスラング等のノイズを手で除去した上で、各単語がレビュー上の表現としてネガティブ・ポジティブ、もしくはどちらでもない、以上 3 つのラベル付けを行った。また、否定助動詞「ない」「ん」「ぬ」「まい」を含む表現をネガティブ表現として扱った。このラベル付けを行ったリスト内に含む単語が、機能要素に属する単語に係った場合、機能要素に付いているラベルの評価を与える。

投稿レビューを基に構築する、機能要素、感情表現リスト及び、提案手法の例を提示する。

例.機能要素

要素①：手・入力・レシート・撮影・カメラ・写真・読み込み・読み取り・精度

要素②：収入・支出・把握・資産・管理・お金・家計・アプリ・口座・登録・メール・クレジットカード・フォワード・マネー・連携・銀行・残高・確認・財布・カード・現金・自動・金融・機関

ラベル付けリスト。

ポジティブ表現：最高・素晴らしい

ネガティブ表現：遅い

以上の機能要素と、ラベル付けリストがあった時に

文書：「アプリのデザインは最高だけれども、入力が遅い」というレビューを取り上げて、評価を行ったとする。

係り受け分解を行うと

「アプリの/デザインは/最高だけれども/入力が/遅い」となる。ここで分解した文節に含まれる単語は要素②に含まれる「アプリ」、要素①に含まれる「入力」の2単語である。

以上2単語を含む文節「アプリの」「入力が」の2つは、それぞれ「デザインは/最高だけれども」「遅い」に掛かっており、要素①はプラスの評価、要素②はマイナスの評価が導出される。

また、提案手法における係り受けのルールは以下の通りで、ネガティブ・ポジティブ表現の評価を行うよう定める。

①：文書中のラベル付けリストの単語から要素に含まれる単語に係った場合、評価を行う。

②：文書中に要素に含まれる単語から、ラベル付けリストの単語に係った場合も①と同様に評価を行う。

③：①、②のルールを踏まえた上で、連続して係り受けが連なった場合も、要素とラベル付けリストの単語に評価を行うものとする。

③の例として、

「各口座の残高だけでなく、証券口座や nanaco カードまでカバーしていて、総合管理できる点が素晴らしいです」という投稿レビューがあったとする。

これを CaboCha によって分解したとき

表 1.連続した係り受けの例

ID	各文節	係り先 ID
0	各口座の	1
1	残高だけでなく	3
2	証券口座や	3
3	nanaco カードまで	4
4	カバーしていて	5
5	総合管理出来る	6
6	点が	7
7	素晴らしいです	終

表 1 のような係り受けの分解に分ける事が出来る。

ここでは要素に属する ID1 の「残高だけでなく」は ID3 の「nanaco カードまで」の文節に掛かるが、ID3 の「nanaco カードまで」は、ID4 の文節へ。

ID4 の文節は ID5 の文節と連なって係っている。

③のルールから、要素に含まれる「残高」は最終的に ID7

の「素晴らしいです」に掛かる事により、ポジティブの評価を受ける事になる。

このように機能要素に対しネガティブ・ポジティブの評価が定まると、それぞれに応じて当該レビューの評価値に-1、あるいは+1の補正を行う、そして最終的にすべてのレビューの評価値の平均を、その機能要素に関するアプリの評価値とする。

4. 提案手法の評価

本研究では、次の2つの実験により提案手法の評価を行った。

実験 1.照らし合わせ実験

投稿レビューから構築した機能要素へのネガティブ・ポジティブの判定に関して、提案手法と、人手による正解データの作成の両方で、ラベル付けを行う。結果の照らし合わせにより、機能要素および、ネガティブ・ポジティブ表現の構築の精度評価。

実験 2.被験者実験

被験者による、アプリケーションのレビュー評価及び、共起ネットワークの有用性評価。また、実験 1 で作成した正解データと被験者のレビュー評価の比較を介した、提案手法によるアプリケーションの評価。

提案手法に用いるデータ及び、実験 1,2 の説明は次節で行う。

4.1 提案手法評価に用いるデータ.

iOS スマートフォンを対象としたアプリケーションの投稿レビューを用いて提案手法の評価を行った。iOS のアプリストア AppStore からレビューを取得するにあたって、API を用いたレビューの取得が可能だが、Apple 社が公開している API の都合上、最新の投稿レビューから少数の件数までしかさかのぼってレビューを取得する事が出来ないのが現状である。そこで、本研究ではアプリケーションのレビュー取得を、アプリケーションのデータ収集を行っている企業 AppAnnie[11]の公開している情報から間接的に取得し、利用する事で提案手法の評価を行った。

今回、用いたスマートフォンアプリケーションは以下の通りである。

- ① 家計簿 Dr.Wallet:レシート撮影で人気の簡単無料家計簿(以降, DrWallet)
- ② 家計簿マネーフォワード-銀行やクレジットカードを連携して自動で支出管理をする無料の人気家計簿アプリ(以降, MoneyFoward)

4.3.1 .アプリケーションの利用方法習得、及び使用.

- 被験者には利用前に、AppStore に記載されているアプリの説明文を読む
- アプリケーションの機能としてチュートリアルが存在する場合、遂行する.
- 上記を行った上で最低でも 30 分間アプリケーションの機能を把握及び習得の為に使用して貰う.
- 最低でも 20 件の領収書をアプリケーションで記録して貰う.その際、アプリケーション内の機能を網羅するように利用する.
- 以上の工程を 5 つのアプリケーション全て行う.

4.3.2 アプリケーションの評価(アンケート).

- 5 つのアプリケーションへの 5 段階評価
- 生成した共起ネットワークから定義した要素 5 つの内、家計簿アプリケーションを利用する上で重視したい項目(複数可)
- 本研究で定義した 5 つの要素について、各アプリケーションの 5 段階評価
- 共起ネットワークを提示した際に、アプリケーションを選定する判断材料になるかどうか

以上の質問項目に対して、被験者に評価して貰った。被験者は大学生～大学院生の計 8 名.いずれの被験者も家計簿のアプリケーションは利用した事がなかった。

被験者実験において、アンケートを作成して回答して貰った.また、利用したアプリケーションの感想が混同しないように、他のアプリケーションをまたぐ利用は少なくとも最低 10 分空けるよう求めた。

5. 実験結果

5.1 照らし合わせ実験

5.2 で設けた規則を基に、ネガティブ・ポジティブ判定の照らし合わせを行う際、表 4 の a~h のいずれかの結果が導出される.導出結果の内、表 4 の a(True Positive)及び e(True Negative)にあたる導出数から、再現度、適合度、F 値を導出した。

表 4.

提案手法	ポジティブ	ネガティブ	判定ナシ
正解			
ポジティブ	a	b	c
ネガティブ	d	e	f
判定ナシ	g	h	

表 5.

各アプリ	再現率	適合率	F値
DrWallet	0.15942	0.2	0.177419
MoneyFow	0.141304	0.52	0.222222
Okanereco	0.191038	0.54	0.28223
ReceReco	0.097368	0.316239	0.148893
Zaim	0.205645	0.56044	0.300885

5.2 被験者実験及び、正解データによる評価補正

被験者実験の結果及び、正解データによる補正の結果は以下のとおりである.正解データの補正は、ネガティブ・ポジティブの判定が認められるレビューに対して、ポジティブならば評価+1、ネガティブならば評価-1 の計算を行うことで、投稿レビューの数段階評価の補正を掛けた.また、同レビューに対してポジティブ・ネガティブの判定が複数起きたとしても、最高±1 より大きく補正は起こらないとした。

表 6.各アプリの評価平均

	補正前	補正後	被験者
DrWallet	3.9234694	3.981845022	4
MoneyFoward	3.904762	4.245614035	2.875
Okanereco	4.586902	5.350720859	2.75
ReceReco	3.670103	3.546709777	3
Zaim	4.210127	4.732853855	3.875

表 7.補正対象に限ったレビューの平均評価

アプリ	補正対象のみの平均スコア		要素の評価平均
	補正前	補正後	被験者
DrWallet	3.774725275	3.791208791	3.525
MoneyFoward	3.87628866	4.039408867	3.175
Okanereco	4.587926509	5.385826772	2.9
ReceReco	3.677419355	3.55721393	2.875
Zaim	4.19121447	4.722797927	3.425

表 8.被験者が重視する要素(人)

機能要素	重視した人数
①	3
②	2
③	2
④	3
⑤	6

5.3 分析

表 6 と表 7 のデータから、レビュー全体の平均スコアと、補正対象のレビューのみの平均スコア内における、補正前後のスコアに関して、被験者がアプリを評価したスコア、およびスコアに基づいたランク付けに近いかを、分析を行った.分析方法としては、スコアに関しては RMSE(Root Mean Square Error)、ランク付けに関してはスピアマンの順位相関係数をそれぞれ導出した.結果は表 9,10 のとおりである。

表 9.スコアの RMSE

	全体	補正対象のみ
補正前と補正後	0.445303565	0.438304701
補正前と被験者	1.000182364	0.962727982
補正後と被験者	1.391220057	1.352560524

表 10.スピアマンの順位相関係数

	全体	補正対象のみ
補正前と補正後	0.9	1
補正前と被験者	-0.2	0.1
補正後と被験者	-0.5	0.1

6. 考察

6.1 照らし合わせ実験

5.1の照らし合わせ実験の結果の表5より、F値の最高値を導出した Zaim アプリでも 0.3 台、最低値を導出した ReceReco では 0.14 台と、精度としては、良いとは言えない結果が得られた。Zaim アプリでは、適合率の値 0.56044 が F 値の大きさの要因に、一方で ReceReco アプリでは、再現率の値 0.097368 が、F 値が小さく導出された要因となっていると考えられる。これらの F 値に寄与しているポジティブ及びネガティブの提案手法と正解データの照らし合わせにおけるマッチング数は以下の表 11,12 となる。

表 11Zaim

提案手法 \ 正解	ポジティブ	ネガティブ	判定ナシ
ポジティブ	91	21	257
ネガティブ	8	11	108
判定ナシ	35	16	

表 12ReceReco

提案手法 \ 正解	ポジティブ	ネガティブ	判定ナシ
ポジティブ	24	17	126
ネガティブ	13	13	187
判定ナシ	30	20	

表 13.Zaim

Zaim	再現率	適合率	F値
ポジティブ	0.246612466	0.679104478	0.361829
ネガティブ	0.086614173	0.229166667	0.125714

表 14.ReceReco

ReceReco	再現率	適合率	F値
ポジティブ	0.143712575	0.358208955	0.205128
ネガティブ	0.061032864	0.26	0.098859

表より、Zaim アプリでは提案手法と正解データのネガティブ・ポジティブ判定において、True Positive にあたるレビ

ューが 91 件と、他のアプリの照らし合わせと比較して、非常に多くのレビューがポジティブ判定として一致したのが考えられる。

ReceReco アプリの方では、照らし合わせの際に、半数近くが False Positive, Negative の判定が出現し、特に False Negative が全体の誤判定の半数以上を占める事により F 値の精度の低さに繋がった要因だと考えられる。

これらは提案手法における、4.1 で構築したポジティブ・ネガティブ表現リスト及び機能要素の構築手法に問題があると考えられる。どちらのリストも、単純に投稿レビューの文書中に出現する頻度が高い単語が登録されるようになっており、文書が意図する内容を頻度で対処出来なかったことがあげられる。特に、ポジティブ・ネガティブ表現リストでは、投稿レビュー文書内に含まれる二重否定の意味を汲んだ機能要素のポジティブ・ネガティブ判定を導出が出来なかった事が確認された。一方で、機能要素リストの方では、要素①内の{クレジットカード},{カード}の名词に代表される表記揺れ等が起きていることが分かる。直接、精度への影響はないものの、同様の意味を持つ単語が複数リストへ登録されることで、他の単語のリストへの登録が阻害されている可能性も考えられる。

以上から、本研究で構築した、ネガティブ・ポジティブ表現及び機能要素のリストは、Zaim アプリの投稿レビューの文書上に出現する単語に近く、ReceReco アプリの投稿レビューの文書に出現する単語に遠かったと考えられる。以上から、ネガティブ・ポジティブ表現及び、機能要素のリスト構築に関して、頻度だけでなく自然言語処理に観点から細かい条件を設ける必要があると考えられる一方で、類似した機能を保有するアプリであっても、投稿レビュー上に出現する名詞において、感情や機能を表現する単語は類似しない可能性があると考えられる。

6.2 被験者実験及び、正解データによる評価補正

5.2における表をグラフにすると図4になる。

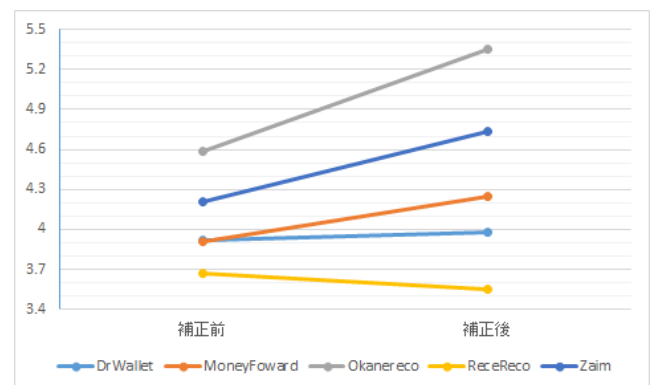


図 4.各アプリにおける補正前及び補正後のスコア

図4より、アプリ ReceRecoを除いた4アプリが補正前に比べて、補正後のスコアが大きくなった。表より、ReceRecoは5アプリの中でも投稿レビューの平均スコアが低かった。

一方で, Okanereco アプリが 5 アプリの中で, 補正後に最もスコアが増加した. これらから, 正解データを用いた機能要素に対する評価の補正は, 元のレビューの評価が高くないアプリは低く, 元のレビューの評価が高い場合は高く補正されることが考えられる. ReceReco アプリでは, ポジティブ表現の単語に比べて, ネガティブ表現となる単語が多く出現し, Okanereco アプリでは全く逆が起きたと考えられる.

6.3 分析

5.3 より, 補正前と補正後における平均スコアと, 被験者実験で得られた平均スコアを用いて RMSE を導出したところ, いずれのアプリにおいても, 補正前に比べて補正後の条件の際に大きい値を取った. また, スピアマンの順位相関係数による導出結果も, 補正後と比較して補正前の方が被験者実験で得られたアプリの評価ランキングに近似していることが分かる. 以上から, 正解データによる補正は, 今回行った被験者実験の結果から遠ざかる結果となった事が言える.

一方で, ランク付けに限っては, 補正対象のレビューのみを対象とした平均スコアによるアプリのランキングは補正前後で変化しなかった. また, 要素①~⑤に限ったアプリの平均スコアを用いて, 同様に RMSE とスピアマンの順位相関係数を導出したところ, RMSE は全アプリで補正前のスコアの方が, 補正後のスコアと比較して, 被験者が評価したスコアに近かったものの, スピアマンの順位相関係数はいずれの機能要素においても, アプリのランク付けが一致しており, 機能要素②に限っては, 補正前に比べて, 補正後の方が被験者のランク付けに近い結果が導出された. このことから正解データから, アプリの評価においてスコア付けの評価においては, 有用性を示せなかったものの, アプリのランク付けに関しては, 有用性を示せたと考えられる.

6.4 今後の課題

将来的な課題としては, 提案手法の精度向上及び, 他アプリへの適用を目標とする.

7. おわりに

本研究では, ユーザがアプリケーション選定において, 判断材料となる情報が少ないという背景から, アプリケーションの機能に関わる単語に注目し, 新たな手法を提案し, 評価を通して, アプリケーションの推薦及び分析を行った.

8. 謝辞

本研究は JSPS 科研費 24300005, 26330081, 26870201 の助成を受けたものです.

本研究を遂行するにあたり, 研究の機会と議論・研鑽の場を提供して頂き, 御指導頂いた国立情報学研究所/東京大学 本位田 真一 教授をはじめ, 活発な議論と貴重な御意見を頂いた研究グループの皆様へ感謝致します.

参考文献

- [1] "KH Coder"
<http://khc.sourceforge.net/>
- [2] "CaboCha"
<https://taku910.github.io/cabocho>
- [3] Bo Pang and Lilian Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, Vol.2, No.1-2, pp. 1-135, 2008.
- [4] Ivan Titov and Ryan McDonald. A joint model of text and aspect ratings for sentiment summarization. In ACL08, pp.308-316.2008.
- [5] 安藤 まや, 関根 聡: "レビューには何が書かれていて, 読み手は何を読んでいるのか?", 言語処理学会 第 20 回年次大会 発表論文集, pp.884-887, 2014.
- [6] A. Begal and T. Zimmermann. Analyze this! 145 questions for data scientists in software engineering. In Proceedings of the 36th International Conference on Software Engineering, pages 12-23, 2014.
- [7] Chen, N., Lin, J., Hoi, S. C. H., Xiao, X. and Zhang, B.: AR-miner: mining informative reviews for developers from mobile app marketplace, Proc. ICSE, pp. 767-778 2014.
- [8] 小川 祐樹, 諏訪 博彦, 山本 仁志, 岡田 勇, 太田 敏澄: "共購買情報を用いた多様性のある推薦アルゴリズムの推薦 -Amazon.co.jp の関連商品ネットワークを用いて-", 日本社会情報学会第 22 回全国大会研究発表論文集 pp.298-303, 2007.
- [9] 花岡 裕, 白井 清昭: "複数の製品を対比する文の検出", 言語処理学会 第 20 回年次大会 発表論文集, pp.896-899, 2014.
- [10] "MeCab"
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [11] "AppAnnie"
<https://www.appannie.com/jp/>