

ユーザ嗜好とレビューとの適合度に基づく モバイルアプリケーション推薦方法の提案と評価

2016SE005 ダンス 満マックスセシヌ 2016SE027 金子 隼介

指導教員 青山 幹雄

1 研究背景と課題

1.1 研究背景

近年、スマートフォンをはじめとしたモバイルデバイスの急速な普及により、モバイルアプリケーション(以下、アプリケーションとする)の利用が増加している。しかし、ユーザは多数存在するアプリケーションの中から自身の要求を満たすアプリケーションを獲得することは困難である。そのため、アプリケーションのユーザの要求を満たすデータを見つけ出し、アプリケーションをユーザに自動で推薦する技術が注目されている[3]。

ユーザへ好ましいユーザ経験(UX)を提供するアプリケーションを推薦するには、ユーザの嗜好に適合するデータを収集し、分析する方法が必要である。そこで、具体的な評価であるレビューを解析することで、アプリケーションに対する要求や利用制約を明確にでき、ユーザのコンテキストに応じた推薦が期待できる。しかし、アプリケーションのレビューは大量にあり、同一のアプリケーションに対する評価がユーザごとに異なる場合がある。そのレビューの中からユーザの嗜好と適合するデータを収集することは困難である。

1.2 研究課題

本研究では以下の2点を研究課題とする。

- (RQ1) ユーザの嗜好に適合するレビューに基づくアプリケーションの推薦方法の提案
- (RQ2) 実際のユーザへアプリケーションを推薦し、提案方法の妥当性評価

2 関連研究

2.1 ペルソナシナリオ法[8]

ペルソナ法とシナリオ法を組み合わせ、ペルソナの視点からシナリオを評価し、ユーザがサービスを利用する際の嗜好や振る舞いを深く理解するユーザモデリング法である。

2.1.1. 構造化シナリオ法[2]

シナリオを段階的に書き分けることで、ペルソナがどのような価値観を持ってどのようにサービスを利用するか記述する方法である。本研究では、利用シーンにおける具体的な振る舞いや感情の変化を記述するインタラクションシナリオをシナリオデータとして解析する。

2.2 推薦システム[5]

アプリケーションのユーザにとって有用と思われる対象、情報、または商品などを選び出し、それらをユーザの目的に合わせた形で提示するシステムである。アプリケーションについての関心や好みの度合いを数値化したデータを推薦システムに入力することでユーザの嗜好を予測する。推薦システムで利用されるスコアとレイティングの定義を以下に示す[1]。

- (1) スコア:アプリケーションに対するユーザの適合度を推薦システムによって定量化したもの。
- (2) レイティング:ユーザがアプリケーションに与える評価。

2.3 ユーザモデリング法を用いた情報推薦システム[3]

ユーザがアイテムのどの要素を重視しているかを推論することによってユーザの価値観をモデリングし、それに基づいてユーザの価値判断を反映させたアイテムを推薦する方法が提案されている。情報推薦における価値観は各要素に対する「こだわり」の強さとして現れると仮定し、“評価抽出”、“ユーザモデリング”、“情報推薦”の3つのモジュールから構成される「こだわり」に着目した情報推薦システムを実装する。

3 アプローチ

ユーザのアプリケーションに対するスコアを計算することでアプリケーションの評価を行う。本研究では、収集できるユーザの嗜好を増やすために、ユーザのコンテキストがレビューに現れやすいモバイルアプリケーションを対象にする。

本研究のアプローチを図1に示す。

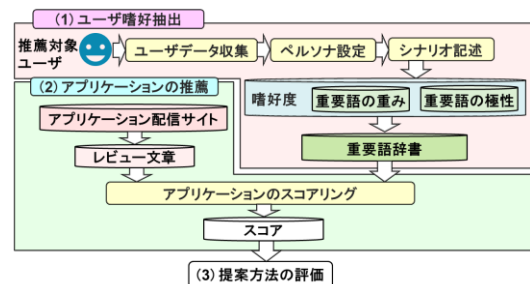


図1 アプローチ

(1) ユーザ嗜好抽出

推薦対象とするユーザのデータを収集し、ユーザに適合するペルソナを設定する。設定したペルソナの振る舞いをシナリオに記述し、シナリオ内からユーザの嗜好に適合する単語を抽出する。抽出した単語を重要語とし、重要語辞書に追加する。このとき、ユーザがどの単語を重視するかを評価するために、重要語の評価極性(1or-1)とシナリオ内の単語の出現頻度に基づく重みを嗜好度とし、重要語に付加する。

(2) アプリケーションの推薦

重要語と一致する単語をアプリケーションのレビュー内から抽出し、重要語のレビュー内での出現頻度と嗜好度に基づいた重回帰分析によってアプリケーションのスコアを計算する。スコアが最も高いアプリケーションが最もユーザに適していると仮定する。

(3) 提案方法の評価

ユーザに推薦候補のアプリケーションを利用してもらい、アンケートによってユーザの各アプリケーションに対するレイティングを収集する。収集したレイティングの順位とスコアリングによる推薦順位を比較し、提案方法の妥当性評価を行う。

4 提案方法

提案方法は以下の3つのプロセスからなる。

- (4.1) ペルソナシナリオ法によるユーザ分析
- (4.2) 重要語辞書の生成
- (4.3) アプリケーションのスコアリング

4.1 ペルソナシナリオ法によるユーザ分析

ユーザ分析プロセスを図 2 に示す。

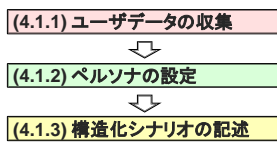


図 2 ペルソナシナリオ法によるユーザ分析
本研究では UX を以下の 3 つの視点から分類する。

- (1) ユーザの嗜好
ユーザの好みによるサービスに対する考えや感じ方。
- (2) 振る舞い
ユーザがどのようにサービスを利用するか。
- (3) コンテキスト
ユーザとサービスを取り巻く環境、制約、条件。

4.1.1. ユーザデータの収集

ユーザデータの収集プロセスの詳細を以下に示す。

- (1) 推薦アプリケーションのカテゴリ決定
アプリケーション配信サイトであらかじめ定義されている利用目的に応じた分類(以下、カテゴリとする)から推薦カテゴリを 1 つ決定する(例: 天気)。
- (2) 仮説に基づいたアンケートの項目設定
決定したアプリケーションのカテゴリに応じたアンケートを作成する。質問項目から得られるユーザの意図やコンテキストを仮説とし、仮説に基づいてアンケートの項目を設定することで、ペルソナの設定とシナリオの記述の方法が明確になる。
- (3) アンケートの実施によるユーザデータ収集
アンケートをユーザに対して実施し、データを収集する。

4.1.2. ペルソナの設定

アンケート結果をもとにペルソナを設定する。アンケートの仮説をもとに、ペルソナの骨格となる要素をスケルトンとして作成する。スケルトンの属性に対して値を記述することで、ペルソナの特徴を定義する。

4.1.3. 構造化シナリオの記述

設定したペルソナの特徴を考慮し、ペルソナがアプリケーションを利用する際の感情や振る舞いを、構造化シナリオを用いて時系列に記述する。シナリオは一文ずつ解析して単位文として用いるため、文の境界を明確にする必要がある。よって、文末は“。”に統一する。

4.2 重要語辞書の生成

構造化シナリオのテキストデータを係り受け解析し、ユーザの嗜好を重要語辞書に追加する。重要語辞書の生成プロセスを図 3 に示す。

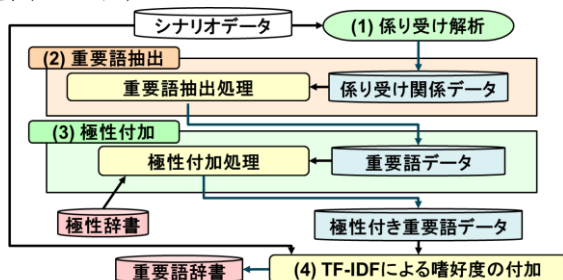


図 3 重要語辞書の生成

重要語辞書の生成プロセスの詳細を以下に示す。

- (1) シナリオの係り受け解析
シナリオデータを係り受け解析器に入力し、係り受け関係データを出力する。
- (2) 重要語抽出
係り受け構造木の深さが 1 以下の名詞を重要語と仮定し、係り受け関係データから重要語を抽出する。
- (3) 極性付加
日本語評価極性辞書[4][6]を用いて重要語データに評価極性を付加する。シナリオは文を単位とした極性と単語の極性が同一の場合が多いため、単位文の極性を単語の極性とする。
- (4) TF-IDF による重みの付加
シナリオデータの TF-IDF に基づく計算を行い、重みを算出する。そして、嗜好度を重要語に付加する。重要語の重みを W 、重要語の極性を P としたとき、嗜好度 D を式(1)で定義する。

$$D = W * P \quad (1)$$

4.3 アプリケーションのスコアリング

アプリケーションのレビューを収集し、重要語とレビューの適合性からスコアを計算する。アプリケーションのスコアリングプロセスを図 4 に示す。

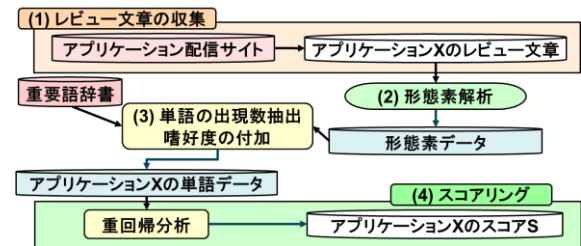


図 4 アプリケーションのスコアリング

スコアリングプロセスの詳細を以下に示す。

- (1) アプリケーションのレビュー文章の収集
決定した推薦カテゴリのアプリケーションレビュー文章をアプリケーション配信サイトから収集する。
- (2) レビュー文章に対する形態素解析
収集したレビュー文章を形態素解析する。
- (3) 単語の出現数抽出と嗜好度の付加
重要語辞書の重要語と一致する単語を形態素データから取り出し、嗜好度を付加する。
- (4) スコアリング
アプリケーションのレビューに含まれる重要語出現頻度と嗜好度からアプリケーションのスコアを目的変数とした重回帰分析を行う。最もスコアが高いアプリケーションがユーザに適していると仮定する。

5 プロトタイプの実装

5.1 実装環境

提案方法のプロトタイプの実装環境を表 1 に示す。

表 1 実装環境

システム	システム名	バージョン
OS	Windows 32ビット版	10
実装言語	Python	3.7.4
形態素解析器	MeCab[7]	0.996
係り受け解析器	CaboCha	0.69
システム辞書	mecab-ipadic	2.7.0

5.2 システム構成

実装したプロトタイプシステムの構成を図 5 に示す。

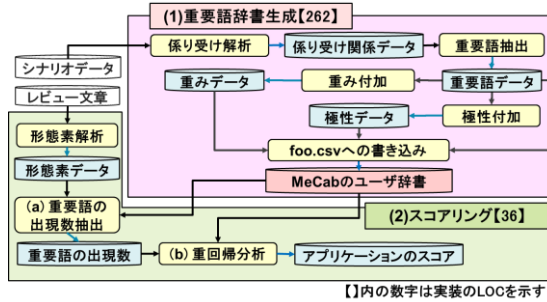


図 5 システム構成

(1) 重要語辞書生成

シナリオデータを解析し、重要語を MeCab のユーザ辞書に定義する処理を実装する。係り受け関係データから重要語を抽出し、foo.csv に記述する。また、重要語ごとの嗜好度を foo.csv の追加エントリに記述する。

(2) スコアリング

推薦対象のアプリケーションのレビュー文章を解析し、スコアを出力する処理を実装する。

(a) 重要語の出現数抽出

システム辞書を用いてレビュー文章を形態素解析する。形態素解析したデータから、MeCab のユーザ辞書に記述済みの重要語データと一致する単語の出現数を抽出する。

(b) 重回帰分析

抽出した重要語の出現数と foo.csv に含まれる重要語の嗜好度から重回帰分析を行い、スコアを計算する。

6 実際のアプリケーションへの適用

6.1 適用方法

ユーザ A, B に提案方法による推薦を行い、アプリケーションのスコアとユーザのレーティングの関係を確認する。この関係から提案方法の妥当性を評価する。また、重要語辞書に含まれる単語がスコアに与える影響を考察する。

6.2 ペルソナシナリオ法によるユーザ分析

“天気予報アプリケーション”を推薦カテゴリとし、アンケートを作成する。天気予報アプリケーションは多くのユーザが“知りたい場所の天気予報を確認する”という目的で利用している。さらに、“雨雲レーダー”や“世界天気予報”などの、ユーザによって利用の有無が分かれる機能が複数あるので、アプリケーション個々の特徴を区別しやすい。また、レビューの数が十分にあるので、評価に適する。また、アプリケーション操作の慣れの差による影響を軽減するために、スマートフォンを普段から利用している人を対象にする。ユーザ A, B にアンケートを実施し、アプリケーションに対する考えを自由記述形式で回答してもらう。アンケートの結果からユーザに適合するペルソナ P_A, P_Bを設定し、ペルソナ P_A, P_Bが天気予報アプリケーションを利用する際の感情や振る舞いを構造化シナリオに記述する。

6.3 重要語辞書の生成

シナリオデータをプロトタイプに入力し、重要語辞書を生成する。ペルソナ P_A の重要語辞書に含まれる単語数と上位下位 5 単語ごとの嗜好度を表 2 に示す。

表 2 ペルソナ P_A の重要語辞書

単語数:44			
上位 5 単語	嗜好度	下位 5 単語	嗜好度
天気	0.105	推移	-0.047
確認	0.091	起動	-0.033
アプリケーション	0.087	不便	-0.013
ため	0.047	面倒	-0.013
ペルソナ	0.043	タップ	-0.013

ユーザの嗜好と明らかに関係のない単語が重要語辞書に追加されていること(“天気”や“アプリケーション”など)が確認できた。ペルソナ P_B の重要語辞書においても同様である。ユーザの嗜好に関係のない単語を重要語辞書から除くために、嗜好度の絶対値に閾値を設定する。閾値は 0.07 から 0.05 まで変化させ、閾値ごとに重要語辞書を生成する。ペルソナ P_A の閾値によって除かれた単語を表 3 に示す。

表 3 ペルソナ P_A の重要語辞書の閾値と除かれた単語

閾値なしから閾値 0.07-0.05 に変更したとき 除かれた単語数:3				
単語名	天気	確認	アプリケーション	
嗜好度	0.105	0.091	0.087	
閾値 0.05 から閾値 0.04 に変更したとき 除かれた単語数:3				
単語名	ため	ペルソナ	推移	
嗜好度	0.047	0.043	-0.047	
閾値 0.04 から閾値 0.03 に変更したとき 除かれた単語数:4				
単語名	表示	カスタマイズ	登録	起動
嗜好度	0.038	0.038	0.033	-0.033

ペルソナ P_A の重要語辞書において、閾値 0.07 から 0.05 の間では取り除かれた単語が同じであることが確認できた。そして、閾値 0.05 から 0.03 に変化する間では、単語数が減少していることから、閾値が 0.05 から 0.03 の範囲は重要語辞書の単語抽出に影響することが確認できた。この重要語辞書の変化がアプリケーションのスコアへ及ぼす影響を確認するため、閾値ごとの重要語辞書を用いてスコアリングを行う。

6.4 アプリケーションのスコアリング

重要語辞書を用いて、アプリケーションごとにスコアを計算する。Google Play Store で「天気」と検索した上位 5 つの天気予報アプリケーションを対象とした。アプリケーションのレビューを関連度順に並べ、各レーティングで約 1,000 文字ずつ、合計約 5,000 文字を人手で収集し、プロトタイプに入力する。ペルソナ P_A の閾値ごとの重要語辞書を用いたアプリケーションごとのスコアリング結果を図 6 に示す。

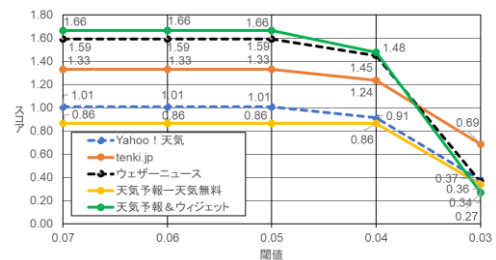


図 6 ペルソナ P_A の重要語辞書を用いたスコアリング結果
このスコアの順位がユーザ A へのアプリケーションの推薦順位となる。図 6 より、閾値が小さくなるほどスコアも低くなっている。特に、閾値 0.04 と 0.03 の間のスコアは閾値 0.05 と 0.04 の間より大きく減少しており、推薦順位も異なっている。これは、正のスコアのうち、重要語“表示”のレビュー内の割合が他の単語に比べて高いからである。ユーザ A のレーティングと推薦順位とのスパイマンの相関係数[1]から、閾値による推薦順位の差がスコアとユーザ A のレーティングとの関係に

与える影響を確認した。

7 提案方法の評価

7.1 評価の目的

提案方法によって計算したスコアとレーティングの間の相関係数が 1.0 に近い値を取れば、ユーザに適合したアプリケーションを推薦できたとと言える。提案方法はアプリケーションの推薦に有効だと言え、妥当だと推定できる。

また、スコアとレーティングの間の相関係数が-1.0 に近い値を取った場合は、スコアとレーティングの間が負の相関となる。この場合、提案方法によって推薦したアプリケーションはユーザに適合していないことが確認できる。しかし、重要語辞書を生成し用いたことで、ユーザのレーティングと相関を持ったスコアを生成できたとと言える。さらに、閾値ごとに相関係数を比較することで、どの閾値が重要語辞書生成に有効か確認する。

7.2 評価方法

推薦候補のアプリケーションをペルソナに適合するユーザに利用してもらい、アンケートを用いてレーティングを収集する。収集したレーティングと提案方法の推薦順位を比較し、相関係数を算出することで提案方法の妥当性を評価する。

7.3 評価用アンケートの実施

評価用アンケートを作成し、ペルソナ P_A, P_B に適合するユーザ A, B に実施する。評価用アンケートを用いて収集したユーザ A のアプリケーションごとのレーティングを表 4 に示す。

表 4 ユーザ A のアプリケーションごとのレーティング順位

アプリケーション名	レーティング	順位
天気予報-天気無料	17	1
天気予報&ウィジェット	16	2
Yahoo! 天気	15	3
ウェザーニュース	12	4
tenki.jp	10	5

7.4 レーティング順位と推薦順位の比較

ユーザ A, B のレーティング順位とペルソナ P_A, P_B に対する推薦順位との相関係数を表 5 に示す。

ユーザ A は他の閾値に比べ、閾値が 0.03 の場合に強い負の相関がある。一方、ユーザ B は閾値 0.03 の場合、他の閾値に比べて相関があるとは言えないが、測定した全ての閾値において-0.6 以下の値となっている。よって、嗜好度の絶対値に閾値 0.03 を設定することで、スコアとレーティングとの相関が高くなる重要語辞書を定義できることが明らかとなった。

また、レーティング順位と推薦順位に強い負の相関があるので、提案方法によりユーザに適合しないアプリケーションが推定できた。しかし、アプリケーションの推薦方法としての妥当性を確認するまでには至っていない。

表 5 ユーザ A, B のレーティングと推薦順位の相関係数比較

閾値	0.07	0.06	0.05	0.04	0.03
ユーザ A の相関係数	-0.3	-0.3	-0.3	-0.3	-0.8
ユーザ B の相関係数	0.1	-0.7	-0.7	-0.6	-0.6

8 考察

8.1 抽出されるユーザの嗜好について

ユーザの嗜好を抽出するために、シナリオ内に含まれる単語を重要語とし、重要語辞書を生成した。従来の推薦技術による嗜好の抽出は、あらかじめ定義されている嗜好を用いるものが多く、抽出できる嗜好の範囲が限定される。そこで、ペ

ルソナのユーザストーリーを係り受け構造木と単語の品詞に着目してユーザの嗜好を抽出した。これより、推薦システム内で定義されていない嗜好を単語として抽出することができたと考えられる。

8.2 嗜好度の閾値について

嗜好度に閾値を設定して、他の単語に比べてスコアへの影響が大きい単語を重要語辞書から取り除いた。ペルソナ P_A の重要語辞書では、閾値を 0.04 から 0.03 に変更したことで、重要語辞書内の単語数が 4 つ減少し、スコアの順位が入れ替わった。その結果、推薦順位とレーティングの間に負の相関が現れた。6.4 章より、重要語辞書内の重要語のうち、嗜好度が 0.03 から 0.04 の単語である“表示”が相関を弱めていることが明らかとなった。現在の重要語辞書を用いると、“表示”などのレビューに多く現れる嗜好度の高い単語の影響が強いスコアが生成されるので、それらの単語を重要語辞書から除去することで相関を強めることができると考えられる。これより、閾値を導入することで、不要な単語の除去が可能であることが確認できた。よって、閾値の導入は重要語辞書の生成に有効である。ペルソナ P_B についても同様となる。

8.3 提案方法の妥当性について

シナリオから抽出した重要語と一致する単語をアプリケーションのレビューから抽出することでユーザのアプリケーションに対する要求やコンテキストを獲得した。推薦順位とレーティング順位に負の相関が現れたことから、ユーザの嗜好に適合しないアプリケーションが推定された。従って、推薦方法としての妥当性の確認には至っていない。しかし、嗜好と適合しないアプリケーションの特徴をレビューから収集することができた。よって、レビューとシナリオの関係を導出できたので、この意味で提案方法は有効である。これにより、嗜好に不適切なアプリケーションを取り除いた推薦が期待できる。

9 今後の課題

- (1) より多くのユーザへ提案方法を適用して、評価を行う。
- (2) 提案方法が有効なレビューの傾向を検証する。
- (3) 類似語や同義語を重要語に含める方法を検討する。

10 まとめ

本研究ではユーザの嗜好に適合するレビューを抽出し、アプリケーションを推薦する方法を提案した。提案方法のプロトタイプを実装し、実際のアプリケーションに適用して妥当性を評価した。重要語辞書とレビュー文章を用いてアプリケーションのスコアリングを行うことで、アプリケーションとユーザの嗜好との相関を明らかにした。

参考文献

- [1] D. K. Agarwal, et al. Statistical Methods for Recommender Systems, Cambridge University Press, 2016.
- [2] 郷 健太郎 他, 構造化シナリオ手法について, 人間工学, Vol. 44 Supplement, Jun. 2008, pp. 38-39.
- [3] 服部 俊一 他, 価値判断に基づくユーザモデリング手法を用いた情報推薦システムの提案とその特性に関する考察, SIG-AM, Vol. 2, No. 3, 2012, pp. 12-18.
- [4] 東山 昌彦 他, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, 2008, pp. 584-587.
- [5] 神島 敏弘, 推薦システムのアルゴリズム(1), 人工知能学会誌, Vol. 22, No. 6, Nov. 2007, pp. 826-837.
- [6] 小林 のぞみ 他, 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, Mar. 2005, pp. 203-222.
- [7] MeCab, <https://taku910.github.io/mecab/>.
- [8] J. Pruitt and T. Adlin, Persona Lifecycle, Morgan Kaufmann, 2006.