

スマートフォンアプリケーションのレビューの自動分類 —自動分類システムの実現のための分類モデルの構築—

2020SE037 宮下拓也 2020SE090 杉本雄大

指導教員：横森励士

1 はじめに

スマートフォンアプリケーション（以下、アプリ）におけるユーザーレビューは、利用者の意見や要望が含まれ、アプリの現状を開発者が知ることができる。開発者はアプリの現状を知ることによって、今後の開発や運営の方向性を定めることができる。Khalid[1]は北米のスマートフォンアプリケーションに投稿された低評価レビューを対象として、レビューに12種類の苦情タイプのタグをつけ、どのような苦情が多いのかなどを調査した。伊藤らは[2]で、投稿されたレビューから単語を抽出し学習モデルを構築し、機械学習に基づいてレビューの自動分類を行う仕組みを提案した。[2]では、[1]の作成した12種類の苦情タイプをもとにして分類していたが、評価実験を行ったあとの考察では、分類モデルに起因する分類精度の低下の原因として「タグ付けの精度における問題」が示されていた。

本研究では、レビュー自動分類システム実現を念頭に置き、レビューの分類のためのモデルを新たに定義し、提案する。新たな分類モデルでは苦情が何に起因しているかに基づいて分類基準を作成した。分類モデルの評価として、作成した分類モデルが分類として妥当であることや判断が正しいことの裏付けとなることを、概要図、Khalidら[1]の分類モデル、利用時の品質モデルと比較することで、分類モデルの категорияに過不足がないことを確認する。加えて、最終的な分類モデルの評価として昨年度のデータ[2]と比較することで、作成した分類モデルの性能を確認する。

2 研究背景

2.1 スマートフォンアプリケーションにおけるレビュー

一般的にアプリはアプリケーションストア（以下、アプリストア）からダウンロード可能であり、アプリストアにはユーザーがタイトル、コメント、星評価などから構成されるアプリのレビューを投稿できる。ユーザーが投稿されたレビューからアプリについての情報を得られるだけでなく、開発者が今後の開発方針を決める指標にもなると考えられ保守、運用を行う際に参考になる情報の1つと言える。

2.2 アプリケーションのレビューを対象とした分析

米国で提供されている無料iOSアプリの低評価レビューについて精査することによって、ユーザーがアプリのどの要素に不満を感じやすいかを調査した Khalidら[1]の研究がある。[1]では低評価レビューにはどのようなコメントが多く存在（苦情頻度）しており、どの要素が低評価に繋がりがやすいか（苦情影響力）を調査するために、レビューを

表1に示す12種類の苦情カテゴリーに分類した。分類した苦情カテゴリーは、保守運用で改善を目的としたリソースの配分時に役立つ情報であると結論づけている。

レビューにおけるタグ付けの自動化を目的として、伊藤ら[2]は機械学習でレビューを分類する方法を提案した。その方法では Khalidら[1]が作成した12種類の苦情タイプに基づいて分類を行っている。[2]の実験では2000件のレビューに対して交差検証を行い、どのくらいの精度が得られたかを紹介している。結果として2000件のレビューに対して交差検証における正解率は約0.68となった。

3 スマートフォンアプリケーションのレビューの自動分類における分類モデルの構築

3.1 過去の分析における問題と研究のアプローチ

伊藤ら[2]の研究では、分類精度の低下の原因として、「タグ付けの精度における問題」が示されていた。考察の結果挙げられる問題として、カテゴリー分けが12種類と多いことや分類が正しいかどうかの判断が難しいことなどが挙げられる。そこで本研究では分類におけるモデルを新たに定義し、その上でタグ付けを行うことを提案する。新たに提案する分類モデルを作るために、将来的なレビュー自動分類システムの観点からステークホルダーを限定した後、苦情に起因する問題を探しだし、それに基づいて分類モデルを作成した。評価では、概要図を用いて苦情に起因する問題を様々なジャンルにおいて網羅できているかを確認することで新しい分類モデルの妥当性を確認する。Khalidら[1]が分類した12種類のカテゴリーや利用時の品質モデル[3]との比較を行い、妥当であることを確認する。最終的な評価として昨年度の実験データに基づいて新しい分類モデルを実験し、昨年度の分類モデルにおける分類精度と比較する。

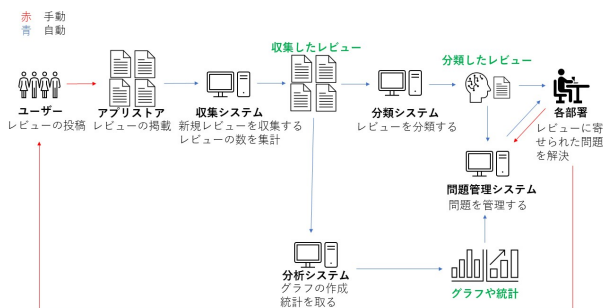


図1 将来的に目指すレビュー分類システム

表 1 Khalid らの分析における苦情の種類 [1]

苦情タイプ	苦情の詳細	レビュー例
強制終了	アプリケーションが強制終了する	起動後、すぐに落ちる
互換性	特定のデバイスや OS のバージョンに問題がある	ipod touch では画面の半分しか見れない
機能削除	特定の機能がアプリケーションを台無しにしている	アプリ自体は素晴らしいが広告を除いてほしい
機能要求	より良くなるために、機能を追加する必要があると感じている	アラートを設定できる機能がない
機能エラー	アプリケーションの特定の問題に言及し、不満を感じている	アプリケーションを開かないと通知が来ない
隠されたコスト	全てを経験するために追加の隠されたコストが必要	リアルマネーを使い、コインの購入を強いてくる
インターフェース設計	デザイン、制御、映像について不満がある	デザインが小奇麗でなく、わかりづらい
ネットワーク問題	ネットワークに問題があるか、応答速度が遅い	新しいバージョンがサーバーにつながらない
プライバシーと倫理	プライバシーを侵す、または反倫理的である	あなたとの接触が目的なアプリケーション
アプリが応答しない	入力の応答が遅い、または全体的に遅い	古いバージョンに戻したい！スクロールが遅い
魅力のない内容	特定のコンテンツが魅力的ではない	画面の見栄えは良いが、退屈つまらないゲーム
重いリソース	アプリケーションがバッテリーまたは容量を消費しすぎる	常時 GPS を使い、バッテリーが消費される
特定できない	ただ単にアプリケーションが悪いと言っている	正直なところ、最悪のアプリケーション

3.2 分類カテゴリーの方針

Khalid ら [1] が示した 12 種類のカテゴリーを用いて実際にタグ付けを行っていくと、タグ付けをする人によって分類が分かれる似たようなカテゴリーを含むレビューが存在することが判明した。そのためタグ付けを行う際、人によって基準が異なるので精度が落ちるのではないかと考えた。また、実際にタグ付けを行った際、Khalid ら [1] が分類した 12 種類の苦情タイプに当てはまらないレビューが存在するのではないかと考えた。

これらの問題を解決するために新しい分類モデルを提案する。作成する分類モデルは、図 1 で示す「レビューを自動化などを通じて簡単に収集し分析することができるレビュー分類システム」を実用化する際に利用したいと考えている。対応すべき会社の部署ごとに考え、苦情がそれぞれの部署に届けられるべきかを考慮する必要がある。そこで各部署の観点からステークホルダーを限定することで対応する問題の範囲を明確にして、苦情に起因する問題に着目し分類する。これによって、似たような種類の苦情をまとめることが容易になり、新たなタイプの苦情があったとしても、誰が対応すべきかという観点からであれば分類しやすいと考えた。これらの理由から新しい分類モデルを作成することで、[2] で問題となっていた「タグ付けの精度における問題」が解決が期待できると考えられる。

4 分類モデルの作成と評価

4.1 分類モデルの作成

将来的なレビュー分類システムを実現するためには各部署ごとに苦情のレビューを届ける必要がある。各部署ごとに送られる苦情のレビューはそれぞれ苦情に起因する問題ごとに分かれていると考えたため、新しい分類モデルでは分類の方針をレビューにおける苦情に起因する問題が何であるかに基づいて分類する。苦情に起因する問題を考えるためにまずは各部署のステークホルダーを限定した。「アプリケーションの部」、「ネットワーク、端末の部」、「お問い合わせ窓口」、「企画部」、「経営戦略部」の 6 つの部をステークホルダーとした。このステークホルダーから苦情に起因

する問題を考えた結果、「アプリケーションの部」は「アプリケーションの問題」、「ユーザーの部」は「ユーザーの問題」、「ネットワーク、端末の部」は「ネットワーク、端末の問題」、「お問い合わせ窓口」は「会社の問題」、「企画部」は「コンテンツの問題」、「経営戦略部」は「ビジネスモデルの問題」となった。この 6 つを苦情に起因する問題として評価前の新しい分類モデルとした。

4.2 評価 1: 概要図上での比較

この新たな分類モデルが苦情に起因する問題を様々なジャンルにおいてすべて網羅できているか確認する必要があると考えた。そこで実際のアプリのレビューを確認し苦情に起因する問題を網羅できているか調査した。調査結果をもとに概要図というアプリ全体の関係から苦情に起因する問題を表している図を作成した。「ショッピングアプリ、オークションアプリ」、「漫画サイト」、「動画配信サイト」、「ゲーム」、「SNS」の 5 つのジャンルで概要図を作成して評価した。図 2 は「ショッピングアプリ、オークションアプリ」の概要図である。簡単な図の説明をする。利害関係者間やアプリケーションに対してどのような関係の要素を持っているかを矢印と要素の名前で表している。例を挙げる。ショッピングアプリに対して落札者はシステム上で落札、支払いなどを行う。ここでは落札と購入を黄色と薄い青色に括っている。理由として落札ではアプリケーションを通じて落札を行うため、アプリケーションの問題や落札などの内容に関するコンテンツの問題が苦情に起因する問題になると考えたためである。このように関係を表す要素に対して色が何色かで括られている場合はその要素が複数の苦情に起因する問題を持っていると考えている。

また新しい問題としてアプリストアの問題が見つかった。苦情に起因する問題を網羅するためにこの問題を新しい分類モデルに加える必要があると考えた。「アプリストアからアップデートできない」などといった苦情レビューが見つかったため、ステークホルダーを外部関係者として「アプリストアの問題」を追加した。その他 4 つの概要図においても「アプリストアの問題」を含めた 7 つの新しい分類モデルで網羅できていたため、将来的に様々なジャン

で使えると評価した。「アプリストアの問題」を追加した7つの苦情に起因する問題を新しい分類モデルとした。表2は新しい分類モデルとその詳細、実際のレビュー例を載せている。

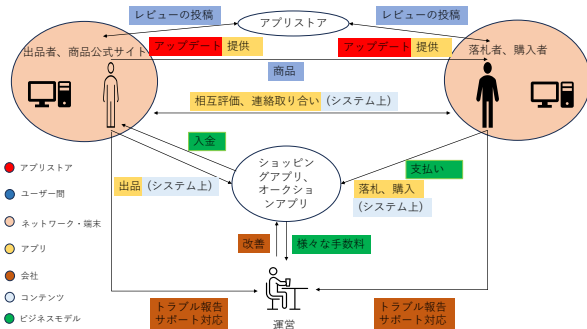


図2 ショッピングアプリ、オークションアプリ概要図

4.3 評価 2:Khalid の分類との比較

新たな分類モデルを利用する際、これまでの苦情の調査で利用してきた Khalid ら [1] が作成した 12 種類のカテゴリーを全て考慮していなければ苦情の種類を全て網羅できているとは言えないと考えられる。表3では今回の分類モデルと [1] の考えた 12 種類のカテゴリーを比較した。例としてビジネスモデルの問題は隠されたコストと対応付けられた。対応付けた理由として、アプリを楽しむための課金や広告視聴は隠されたコストにあたり、これは会社のお金の稼ぎ方に関わってくる問題であると考えたからである。

比較した結果、[1] が作成した 12 種類の苦情カテゴリーは考慮できていた。その上で [1] が分類したカテゴリーには存在していなかったカテゴリーとしてアプリストアに起因する問題があげられた。加えて、[1] の分類モデルのカテゴリーには新しい分類モデル一つに対して複数のカテゴリーで対応するものがあると考えた。

4.4 評価 3: 利用時の品質モデルとの比較

保守をする際に、新しい分類モデルによって分類された苦情を改善することで製品をより良くできるのかを検討する必要がある。そこで、保守をする際に必要となる要素を含んでいる ISO/IEC 25010 利用時の品質モデル [3] に着目した。本項目では、新しい分類モデルで利用時の品質モデルの項目で考慮されていない要素が無いかを確認するため、ISO/IEC 25010 利用時の品質モデル [3] と新しい分類モデルを比較する。図3は利用時の品質モデルと比較した図である。アプリケーションの問題では、この問題には有効性と満足性の副特性である信用性と快適性が当てはまると考えられる。アプリケーションの問題における快適性を例とすると、「アプリケーションを利用する際に、ユーザーインターフェースが利用しやすいか。」などがあげられる。

比較の結果として、新たな分類モデルは品質モデル上で全て対応していると考えられる。会社の問題は運営と利

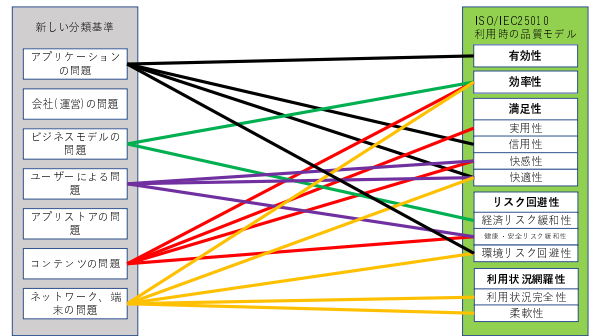


図3 ISO/IEC 25010 利用時の品質モデルとの比較

用者の関係であり、アプリストアの問題は利用者が「App Store」や「Google Play」を使う。これらの問題ではシステムの利用を想定していないため利用時に直接関わっていないと考えたため対応していないと考えた。

4.5 評価 4: 昨年度との分類精度における比較

実用化に向けてタグ付けのしやすさだけではなく機械学習の分類精度を上げる必要がある。そのため 10 分割交差検証において新しく作成した分類モデルを使った場合の機械学習の分類精度を求めて昨年度の機械学習の分類精度と比較する必要があると考えた。

昨年度 [2] 使用した 2000 件のレビューを今回の新しい分類モデルでもう一度タグ付けした。まずは 10 分割交差検証を行った。最初の 100 件で 10 分割交差検証した際の機械学習の精度が平均で 0.61999 となった。200 件の機械学習の精度は平均で 0.7 であった。1500 件の 0.726 が全体において最高の数値となった。その後、0.7 を前後するという結果になった。最終的に 2000 件の機械学習の精度は 0.7105 となった。図4は昨年度と今年度の 10 分割交差検証における機械学習の精度を比較した図である。結果として昨年度よりも機械学習の精度が上がった。

5 考察と今後の課題

今回作成した分類モデルを様々な視点からどのような点から有益であるかを考えた。

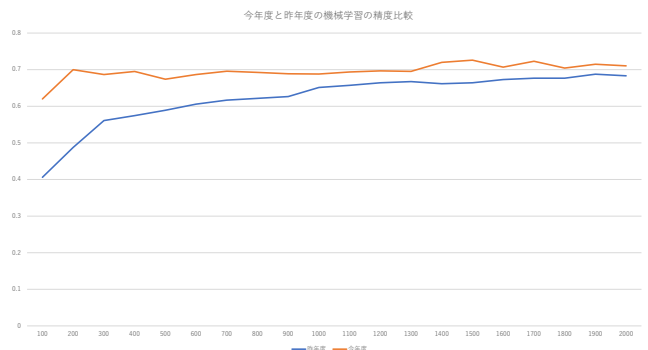


図4 10分割交差検証における機械学習の精度比較図

表 2 新たな分類モデル

分類モデル	苦情の詳細	レビュー例
アプリケーションの問題	アプリケーションの制御事態が原因となって引き起こされる問題	アプリが強制終了する、操作しづらい
会社の問題	会社の運営の仕方が問題となって引き起こされる問題	ユーザーの問い合わせに答えられない
ビジネスモデルの問題	企業のお金の稼ぎ方が原因となって引き起こされる問題	広告が多い、課金システムの問題
ユーザーによる問題	悪質なユーザーが原因となって引き起こされる問題	マナーの悪いユーザーがいる
アプリストアの問題	アプリストア自体が原因となって引き起こされている問題	アップデートができない
コンテンツの問題	アプリの内容自体が原因となって引き起こされる問題	面白くない、他のアプリの方が面白い
ネットワーク、端末の問題	利用者の利用環境が原因となって引き起こされる問題	特定の端末、ネットワークでは動かない

表 3 Khalid の分類との比較

起因する問題	12 種類の苦情タイプ
アプリケーションの問題	強制終了・互換性・機能削除 機能エラー・インターフェイス設計 ネットワーク問題・重いリソース アプリケーションが応答しない
会社の問題	機能要求
ビジネスモデルの問題	隠されたコスト
ユーザによる問題	プライバシーと倫理
アプリストアの問題	対応無し
コンテンツの問題	プライバシーと倫理・魅力のない内容
ネットワーク、端末の問題	強制終了・互換性・ネットワーク問題 アプリケーションが応答しない 重いリソース

- 分類先が集約されたこと
本研究で新しい分類モデルを 7 つに集約した。昨年度の課題でもあった「タグ付けのしやすさ」を改善できたと考える。またステークホルダーを限定し、苦情に起因する問題に着目することで将来的な自動分類システムの実現の際に誰が対応すべき問題かが分かるようになったことがよかったことであると考察した。
- Khalid が作成した分類モデルを考慮できていたこと
新しく作成した分類モデルが Khalid が作成した分類モデルを考慮できていなければ苦情の種類を全て網羅できていたとは言えないため、考慮できていたことがよかったことであると考察した。
- 利用時の品質モデルを考慮できていたこと
新しく作成した分類モデルが ISO/IEC 25010 利用時の品質モデルを考慮できていたため、今回作成した分類モデルは保守の観点で必要な要素が揃っていると考えられる。このことから、この分類モデルを利用して分類を行うことが製品をより良くすることに繋がることが良かったことであると考察した。
- 昨年度より機械学習の精度が高かったこと
10 分割交差検証において今年度は最終的に 0.7105 という機械学習の精度になった。この結果は昨年度の結果よりも高いため良かったことであると考察した。

5.1 今後の課題

改善すべき問題として、複数タグの問題はまだ解決していない。Khalid が作成した分類モデルとの比較において、新しい分類モデルの一つのカテゴリに対して Khalid が

作成した分類モデルの複数に対応していた。本分類でも、実際のレビューに複数のタグが付くであろうコメントが多く複数のタグに関連する問題がある。改善策として複数のタグに当てはまるか判別できる学習モデルを作成することがあげられる。

また、レビューのコメントの冗長さによる問題が存在する。機械学習において不正解になったレビューの中には苦情のコメントに続いて感想であったり関係のないコメントが多く見られた。「本当に面白い！ただ、いくら 0 円でも最終話辺りは課金しなきゃ行けない様子なのが悲しい。」が実際のレビュー例である。このような苦情に関係のないコメントに機械学習が誤って分類している可能性があることと考察した。苦情に関係のないコメントを削るなどデータを加工していくことが必要である。

6 おわりに

本研究では、レビューの自動分類システムの実現のため、苦情に起因する問題の観点から新しい分類モデルを提案した。その分類モデルを、作成した概要図、Khalid[1] の分類モデル、利用時の品質モデル [3]、昨年度の研究データ [2] との比較を通してそのモデルの妥当性について確認した。今後の課題として、レビューの自動分類システムの実現に加えて、精度向上のための仕組みをどのように用意していくかがあげられる。

参考文献

- [1] Hammad Khalid, Emad Shihab, Meiyappan Nagappan, Ahmed E. Hassan : “What Do Mobile App Users Complain About?”, In IEEE Software, Vol.32, No.3, pp.70-77, 2015.
- [2] 伊藤陽, 紀本光琉 : “スマートフォンアプリケーションのレビューにおける苦情の分析—レビューの自動分類に関する考察—”, 南山大学理工学部, 2022 年度卒業論文, 2023.
- [3] JIS X 25010:2013 (ISO/IEC 25010:2011) システム及びソフトウェア製品の品質要求及び評価 (SQuaRE) — システム及びソフトウェア品質モデル, 日本工業規格, 2013