

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

—地域による傾向の違いの調査—

2015SE047 松永夏季

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）を開発する際、ユーザーが投稿するレビューは開発者にとって、ユーザーからの重要なフィードバックとみなすことができる。本研究では、過去に行われたレビュー中の苦情内容の分析を踏まえ、米国のアプリにおいて高評価レビューでどのように苦情が存在するか、他の地域（英国）での苦情内容の分析を行う。「ユーザーレビューを地域ごとに得た場合、共通点と相違点が存在し、共通点は全ての地域においてアプリ開発・保守をする上で気を付けるべき点、相違点は各地域ごとに気を付けるべき点を表す。」という仮説を立て、レビュー内容を分析し、地域性を保守にどのようにフィードバックを行うべきかを考察する。

2 レビューにおける苦情内容の分析

Khalid ら [1] は、米国で提供されている無料 iOS アプリを対象として、低評価レビューの内容を分類した。結果として、レビューは 12 種類の苦情タイプに分類され、機能に問題が生じている「機能エラー」と機能を求める「機能要求」が苦情に多く存在することを示した。安部ら [2] は、日本のアプリに対して条件をそろえたうえで低評価レビューの内容を分類した。結果として、「機能エラー」などは同じくらい存在したが「機能要求」は少なく、低評価のユーザーが消極的であることがわかった。平井ら [3] は、日本のアプリでは、中・高評価のレビューに提言となる「機能要求」が多く存在することを示した。

本研究では、米国のアプリでも同等の傾向がみられるかを確認する。他の地域（英国 iOS アプリ）でも同様の項目がみられるかについて調査し、各国のレビューに地域差が存在することを示す。調査項目を以下のように設定した。

1. 米国・英国のアプリでの、中・高評価レビューで苦情が存在した割合。日本と同様に、中・高評価レビューでも不満が存在することを示す。
2. 米国のアプリのレビュー全体における各苦情内容の割合。低評価だけの場合と比較し、違いを調査する。
3. 英国のアプリのレビューにおける各苦情内容の割合。日本・英国・米国の結果の共通点と相違点を評価する。

3 実験

米国、英国の分析対象のアプリについて、過去の研究とできるだけ同じ条件となるよう表 1 のように選びなおした。約 2 ヶ月分のレビューを入手し、信頼水準 95 %、信頼

区間 5 % でアプリ毎にレビューの抽出件数を決定し、苦情の内容を分類した。表 2 は日本・米国・英国のアプリのレビューにおいて、苦情が出現した割合を示す。どの地域でも、ほぼ全ての低評価レビュー、大部分の中評価レビュー、一部の高評価レビューに苦情が含まれており、各評価帯における苦情の内容調査する価値があることがわかる。

表 1 抽出対象となるアプリケーション

	アプリケーション	ジャンル	サンプリング	低評価 アプリケーション数	高評価 アプリケーション数
米国のアプリ [1]	20 個	15	264~383	10 個	10 個
日本のアプリ [3]	20 個	13	255~375	10 個	10 個
米国のアプリ	20 個	13	222~367	10 個	10 個
英国のアプリ	20 個	13	79~293	10 個	10 個

表 2 レビューにおける苦情の出現割合

	星 1, 2	星 3	星 4, 5	計
米国のアプリ	0.99 (3288/3289)	0.97 (638/656)	0.42 (766/1845)	0.81 (4692/5790)
日本のアプリ [3]	1 (3843/3843)	0.85 (603/713)	0.37 (834/2267)	0.77 (5280/6823)
英国のアプリ	0.99 (1408/1410)	0.99 (256/258)	0.25 (463/1834)	0.61 (2127/3502)

3.1 レビューにおける苦情内容の分類

表 3 は、苦情を含むレビューを母集合とした場合の、各苦情タイプが占める割合を示す。先行研究 [1] での結果と、今回米国のアプリに対して集計した結果をそれぞれ比較している。表では、苦情タイプ毎に 20 個のアプリそれぞれでの出現割合の中央値を求め、多い順番に並べた。低評価同士で比較した場合、占有率の基本的な傾向は大きく変わっていない。[1] では 4 位であった「ネットワーク問題」の順位が下がっており、ネットワーク環境が改善していることが分かる。また [1] では 10 位であった「アプリが応答しない」の順位が上がっており、安定性が重視されている。表 4 は、米国のアプリについて、各評価帯における苦情を母集合とした場合の、各苦情タイプが占める割合を示す。中央値が 0 の項目については、平均値で順位付けを行っている。高評価になるほど、「機能要求」や「機能削除」の割合が高くなり、積極的に利用者が開発者に向けて提言していることがわかる。表 5 は、英国のアプリについて、各評価帯における苦情を母集合とした場合の、各苦情タイプが占める割合を示す。英国と日本 [3] では、低評価レビューにおける「機能要求」の出現頻度が米国と比べて低いため、英国と日本のアプリの利用者はあまり提言には熱心ではないと考えた。共通点として、「機能エラー」、「強制終了」、

「アプリが応答しない」の割合が低評価レビューに多く見られた。評価が上がるごとに「機能要求」の割合が上昇しており、アプリの改善点を具体的なコメントで残していた。

表3 各苦情タイプの苦情の出現頻度 (米国のアプリ)

苦情タイプ	星1~5(占有率)		星1, 2		星1, 2[1]	
	順位	中央値 (%)	順位	中央値	順位	中央値
機能エラー	1	31.24	1	29.94	1	26.68
機能要求	2	21.35	2	11.89	2	15.13
強制終了	3	6.35	3	7.09	3	10.51
インターフェース設計	4	3.74	5	4.08	5	3.44
アプリが応答しない	5	3.63	4	4.73	10	0.73
プライバシーと倫理	6	3.21	6	2.78	9	1.19
機能削除	7	2.73	10	1.88	6	2.73
ネットワーク問題	8	2.08	8	2.27	4	7.39
魅力のない内容	9	1.97	7	2.4	11	0.29
互換性	10	1.79	9	1.94	8	0.39
隠されたコスト	11	1.404	12	0.75	7	1.54
重いリソース	12	1.403	11	1.44	12	0.28
特定できない	-	2.95	-	3.41	-	13.28

表4 各評価帯での苦情の占有率 (米国)

苦情タイプ	星1, 2		星3		星4, 5	
	順位	中央値 (%)	順位	中央値	順位	中央値
機能エラー	1	29.94	1	30.56	2	21.95
機能要求	2	11.89	2	29.71	1	50.5
強制終了	3	7.09	3	5.16	5	(7.29)
アプリが応答しない	4	4.73	6	2.18	8	0(1.92)
インターフェース設計	5	4.08	4	4.94	4	2.26
プライバシーと倫理	6	2.78	11	0(1.67)	10	0(1.31)
魅力のない内容	7	2.4	12	0(1.38)	12	0(0.44)
ネットワーク問題	8	2.27	9	0(2.52)	11	0(0.71)
互換性	9	1.94	10	0(1.84)	6	0(3.53)
機能削除	10	1.88	5	2.41	3	2.73
重いリソース	11	1.44	7	1.76	9	0(1.36)
隠されたコスト	12	0.75	8	0(3.08)	7	0(2.41)
特定できない	-	3.41	-	1.11	-	0(3.61)

表5 各評価帯での苦情の占有率 (英国)

苦情タイプ	星1, 2		星3		星4, 5	
	順位	中央値 (%)	順位	中央値	順位	中央値
機能エラー	1	38.89	1	27.43	1	38.18
強制終了	2	9.43	4	0(9.50)	2	0.89
機能要求	3	7.33	2	21.11	3	25.54
アプリが応答しない	4	6.75	5	0(3.38)	6	0(4.01)
プライバシーと倫理	5	3.91	12	0(0.63)	10	0(0.54)
インターフェース設計	6	3.32	3	5.24	4	0(6.16)
魅力のない内容	7	1.67	11	0(1.03)	-	0(0)
互換性	8	1.32	8	0(1.95)	5	0(4.46)
ネットワーク問題	9	0.67	6	0(3.35)	9	0(0.80)
重いリソース	10	0.5	7	0(2.92)	8	0(3.55)
隠されたコスト	11	0.39	10	0(1.25)	11	0(0.43)
機能削除	12	0(1.30)	9	0(1.53)	7	0(3.92)
特定できない	-	9.5	-	2.38	-	0(1.52)

4 考察

本研究では「ネットワーク問題」に関する苦情の出現割合が [1] と比べて低く、Khalid ら [1] の研究が行われた 4 年前と比較し、インターネット環境が向上したと考えられる。利用者が良いアプリと判断する基準は高くなっている

とも考えられる。さらに、英国・米国・日本の 3 か国とも「アプリが応答しない」、「機能エラー」、「強制終了」の出現頻度が高く、アプリの安定性についての苦情タイプに厳しい意見が多く見られた。多くのアプリが存在する現在の環境において、他のアプリと区別されるためにはこれらアプリの安定性の確保が最低条件であることがわかる。

米国の中・高評価のレビューでは、“現状はアプリの内容に満足しているが、このような機能を追加してほしい”などの「機能要求」や、“このような機能が無くなればアプリがより良くなる”といった「機能削除」の苦情が多く見られた。一方で、アプリに対しての不満だけが述べられている苦情は、高評価レビューにはほぼ存在しなかった。低評価レビューにおいて、日本 [3] では「魅力のない内容」の苦情タイプの出現割合が 3 位、英国では単に“悪い”などといった苦情タイプの特定ができない具体性のないレビューが全体の約 10 % と多く出現したが、米国のアプリの低評価レビューでは、低評価に至った具体的な理由がコメントされているものが多かった。米国では、日本や英国と比べると、ユーザーが自身の持つ意見を具体的にコメントに残し、低・中・高評価に至る理由を明確に表していた。

米国と比較し英国と日本 [3] では、各評価帯全てにおいて「インターフェース設計」の苦情タイプの平均値が高い結果となった。さらに、占有率の値より平均値が高くなっている。このことから、一部のアプリでは、「インターフェース設計」が評価を左右している場合があることがわかる。日本と英国では米国よりも、アプリの操作性やデザインについて考慮する必要があるといえる。

5 まとめと今後

本研究では、米国と英国のレビュー全体における苦情傾向を調査し比較を行った。米国のユーザーは提言に熱心である一方で、日本や英国のユーザーはそれほど熱心ではないなど、地域性を確認した。苦情傾向の相違点を考慮し、改善の方向性を決定する必要があることがわかる。今後、アップデートが原因の苦情が多数あったことから、アプリの低評価とアップデートの関係を調査したい。

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] 安部寛生, 波多野雅信, 小林佑汰 : “日本のスマートフォンアプリケーションにおける評価の低いユーザーレビューでの苦情内容の分析”, 南山大学理工学部 2017 年度卒業論文, 2018.
- [3] 平井賢人, 稲垣絢也 : “スマートフォンアプリケーションにおけるユーザーレビューの内容の分析—低評価レビューと高評価レビューの傾向の違いについて—”, 南山大学理工学部 2018 年度卒業論文, 2019.