

On the Violation of Honesty in Mobile Apps: Automated Detection and Categories

Humphrey O. Obie, Idowu Ileku, Hung Du, Mojtaba Shahin, John Grundy, Li Li, Jon Whittle, Burak Turhan

発表者 花田

背景

- 人間的価値
 - 人々が人生において何を重要視するかの指針となる原則
→誠実さ, プライバシー, 好奇心, セキュリティ, 正直さなど
- この人間的価値は, モバイルソフトウェアアプリケーション(アプリ)によって侵害される可能性がある
 - 悪影響は社会的に様々な形で現れる可能性
- 本研究では, 誠実さと侵害に注目する

目的

- レビューにおける誠実さという人間的価値の侵害の自動検出を支援するため
 - 機械学習モデルとデータセットを提示
- アプリによる誠実さの侵害の自動検出だけでなく
 - ユーザーがアプリのレビューで表現した誠実さの侵害のさまざまなカテゴリを理解

※レビューは, ユーザーの課題や要求を検出するための代理[5, 14, 22, 46, 62]

[5] Eman Abdullah AlOmar, Wajdi Aljedaani, Murtaza Tamjeed, Mohamed WiemMkaouer, and Yasmine N. El-Ghaly. 2021. Finding the Needle in a Haystack: On the Automatic Identification of Accessibility User Reviews. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 387, 15 pages.

[14] Andrea Di Sorbo, Sebastiano Panichella, Carol V. Alexandru, Junji Shimagaki, Corrado A. Visaggio, Gerardo Canfora, and Harald C. Gall. 2016. What Would Users Change in My App? Summarizing App Reviews for Recommending Software Changes. In Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (Seattle, WA, USA) (FSE2016). Association for Computing Machinery, New York, NY, USA, 499–510.

[22] Emitza Guzman and Walid Maalej. 2014. How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews. In 2014 IEEE 22nd International Requirements Engineering Conference (RE). 153–162. <https://doi.org/10.1109/RE.2014.6912257>

[46] Humphrey O. Obie, Waqar Hussain, Xin. Xia, John Grundy, Li Li, Burak Turhan, Jon Whittle, and Mojtaba Shahin. 2021. A First Look at Human Values-Violation in App Reviews. In 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). 29–38.

[46] Humphrey O. Obie, Waqar Hussain, Xin. Xia, John Grundy, Li Li, Burak Turhan, Jon Whittle, and Mojtaba Shahin. 2021. A First Look at Human Values-Violation in App Reviews. In 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). 29–38.

データセット

Table 1: Statistics of the dataset.

Number of Apps	713
App Categories	25
All Reviews	236,660
Honesty-related Reviews (after keywords filter)	4,885
Honesty Violation Reviews (after manual validation)	401

- All Reviews
Elerら[18]の公開データセット + Obieら[46]の公開データセット = 合計236,660件
- Honesty-related Reviews
Obieら[46]が作成した人間の価値観の辞書をベースに、フィルタリング
- Honesty Violation Reviews
二つの研究を基準に、手作業でラベリング + フィルタリング
[8] [40]

Honesty Violation Reviews 4 0 1 件

+

Honesty-related Review 4 0 1 件

↓

8 0 2 件のデータセット

[8] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and inter-coder reliability and agreement. *Sociological Methods & Research* 42, 3 (2013), 294–320.

[18] Marcelo Medeiros Eler, Leandro Orlandin, and Alberto Dumont Alves Oliveira. 2019. Do Android App Users Care about Accessibility? An Analysis of User Reviews on the Google Play Store. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems (Vitória, Espírito Santo, Brazil) (IHC '19)*. Association for Computing Machinery, New York, NY, USA, Article 23, 11 pages.

[40] Elizabeth R Morrissey. 1974. Sources of error in the coding of questionnaire data. *Sociological Methods & Research* 3, 2 (1974), 209–232.

[46] Humphrey O. Obie, Waqar Hussain, Xin. Xia, John Grundy, Li Li, Burak Turhan, Jon Whittle, and Mojtaba Shahin. 2021. A First Look at Human Values-Violation in App Reviews. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. 29–38.

RQ

RQ1.

誠実さを侵害する行為を文書化したレビューを自動的かつ効果的に特定することは可能か？

RQ2.

アプリレビューでは,どのような種類の誠実さへの侵害が報告されているか？

RQ1アプローチ

Table 2: Comparison of confusion matrix and Matthews correlation coefficient (MCC) of classification models.

	SVM	LR	NN	RF	GBT
True negative	0.432	0.407	0.358	0.371	0.358
True positive	0.457	0.469	0.482	0.420	0.420
False positive	0.025	0.049	0.099	0.085	0.099
False negative	0.086	0.074	0.062	0.124	0.124
MCC	0.785	0.753	0.676	0.581	0.555

+

Table 3: Comparison of classification models.

	SVM	LR	NN	RF	GBT
Accuracy	0.889	0.877	0.840	0.790	0.778
Precision	0.949	0.905	0.830	0.829	0.810
Recall	0.841	0.864	0.886	0.773	0.773
F1 score	0.892	0.884	0.857	0.800	0.791

マシューズ相関係数(MCC)を使用

→ Accuracy (正確度)
Precision (適合率)
Recall (再現率)
F1 score (F1値) を計算

5つの分類モデルを使用

- SVM
- LR (ロジスティック回帰)
- NN (ニューラルネットワーク)
- DT(決定木)
- GBT (勾配ブースティング木)

※ 10分割交差検証

RQ1結果

- 誠実さを侵害する行為を文書化したレビューを自動的かつ効果的に特定することは可能か？

SVMがベースラインのランダム分類器の性能を凌駕している
→特定可能かつ性能も良い

Table 4: Comparison of our model to a baseline classifier.

	Our (SVM) approach			Random classifier		
	Precision	Recall	F1	Precision	Recall	F1
Classification	0.949	0.841	0.892	0.0017	0.5	0.0034
Improvement	-	-	-	558.235x	1.682x	262.353x

RQ2アプローチ

- データセットのHonesty Violation Reviews（401件）に対して
→ オープンコーディングの手順 [20] を適用
- 401 件のレビューすべてにラベル付け
→ 10 種類の誠実さへの侵害を特定
 - この 10 種類のカテゴリーに分類
 - ※ カテゴリー分類に対する意見の相違は, negotiated agreement approach [8, 40]を用いて解決

[20] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory; strategies for qualitative research. Nursing research 17, 4 (1968), 364.

RQ2結果

- アプリレビューでは,どのような種類の誠実さへの侵害が報告されているのでしょうか?

- 不当なキャンセル,返金ポリシー
- 虚偽広告
- 欺瞞的購読
- 不正なシステム
- 不正確な情報
- 不正な料金
- サービスなし
- レビュー削除
- なりすまし
- 不正に見えるアプリ

→誠実さ侵害は,10カテゴリーに特徴づけることができる

Table 5: Frequency (f) of app reviews in the honesty violation categories (out of 401 total *honesty_violations* reviews – note that some reviews fall into multiple categories).

	Unfair cancellation and refund policies	False advertisements	Delusive subscriptions	Cheating systems	Inaccurate information	Unfair fees	No service	Deletion of reviews	Impersonation	Fraudulent-looking apps
f	48 (12%)	55 (14%)	33 (8%)	93 (23%)	15 (4%)	106 (26%)	64 (16%)	6 (1.5%)	9 (2%)	29 (7%)

議論と有効性への脅威

- 誠実さという価値は社会で倫理的に望ましい価値
 - モバイルアプリを含む誠実なソフトウェア成果物の開発の促進が必要
- アプリにおける購読プロセスの非透明性
 - アプリの解約, リマインダー機能が必要
- SEにおける人間の価値観の研究は, まだ初期段
 - 研究を進めることで, 人間の価値から恩恵を受けれるのでは？

有効性への脅威

- 内部妥当性 • 構成妥当性 • 外部妥当性

結論

- 誠実さという人間の価値観の違反を明らかにするアプリレビューを自動で見つけるためのアプローチを提示
- 検証したレビューのデータセットを用い,5つの異なるアルゴリズムを評価
 - その結果,SVMアルゴリズムが,高い精度でレビューの誠実さの侵害を検出できることがわかった
- アプリレビューに反映される誠実さの侵害の種類を特徴付けをおこなった
- 誠実さに反するレビューを手動で定性的に分析した結果,10のカテゴリに分類した

所感

- 誠実さという人間の考えを定量化することの難しさがわかった
- 8ヶ月ものラベリング,フィルタリングを行ったことへの驚き