

ソフトウェアデータ収集におけるユーザの態度に影響する 要因の分析

村上優佳紗^{†1,a)} 高塚由利子^{†2} 角田雅照^{†2,b)}

概要: ソフトウェアの品質改善やソフトウェア開発プロセスの改善のためには、分析のためのデータを収集することが重要となる。近年、このデータをユーザが使用しているコンピュータから収集し、サーバーに送信することが行われている。例えば Android OS 端末では、ソフトウェアのクラッシュ時に、クラッシュ時のソフトウェアの情報をユーザが入力したコメントとともにソフトウェアの開発元に送信するフォームが現れる。本稿では、このようなユーザからのデータ収集を促進するために、ユーザのデータ収集に対する協力態度に影響する要因を分析する。収集対象のデータとして、クラッシュレポート、アプリケーションの操作ログ、ソフトウェア開発の操作ログの3種類を想定し、21人の情報科学を専攻する学生にアンケートを行った。アンケートでは、各データ収集に対する協力姿勢と、協力姿勢の理由を選択してもらった。さらに、データ収集元の信頼度が高いと仮定した場合の協力姿勢の変化についても確かめた。その結果、データ収集へのユーザの協力姿勢を改善するためには、データ収集に協力した場合のメリットを明確に説明する必要があることと、データの収集元へのユーザの信頼度を高める必要があることがわかった。

キーワード: データ収集, 信頼, リスク, アンケート

Analysis of Factors Affecting Users' Attitude on Software Data Collection

YUKASA MURAKAMI^{†1, a)} YURIKO TAKATSUKA^{†2} MASATERU TSUNODA^{†1, b)}

Abstract: Collecting data relating to software is essential to improve software quality and software development process. In recent years, such data is sometimes collected on users' computer, and it is sent to server computer. For example, Android OS shows a form which sends data of software and comments manually input by users, when the software crashes. To promote the data collection, this paper analyzes factors which affects users' attitude to the data collection. We assumed three types of data collection, i.e., crash report of software, user operation of applications, and log of users' activity on software development. We asked 21 students whose major was information science. In the questionnaire, we asked user's attitude (e.g., cooperative or not) to the data collection, and the reason of the attitude. Additionally, we asked user's attitude, assuming the trust to an organization which collects the data was high. The result suggested that benefit of the cooperation should be explicitly explained, to enhance intent of the cooperation. Also, trust to the organization which collects the data was also important to the cooperation.

Keywords: Data collection, trust, risk, questionnaire

1. はじめに

近年、ソフトウェアに関するデータをユーザや開発者から収集するという試みが多くなされている。バグに関するものでは、例えば携帯情報端末向けの OS である Android 上でアプリケーションがクラッシュした場合、図 1(a)のようなメッセージダイアログが表示される。ここで「フィードバックを送信」を選択すると、(b)のようなフォームが表示され、システムログを合わせて送信することができる。macOS でも同様の機能があり、Microsoft Windows 7/8/10 などでも、問題が発生したときに、問題レポートを送信するかどうかをユーザが選択することができる。

個々のアプリケーションが同様の機能を持つ場合もあ

る。例えば Web ブラウザの Mozilla Firefox の場合、ソフトウェアがクラッシュすると、クラッシュレポートを送信するためのフォームが表示され、ユーザがレポートを送信するかどうかを選択することができる。

バグ以外のデータをユーザから自動で収集する場合もある。例えば Windows の場合、「カスタマーエクスペリエンス向上プログラム」にユーザが同意すると、ユーザのシステム構成や特定の種類のイベントに関する統計情報が収集され、定期的送信される。これにより、オペレーティングシステムの信頼性やパフォーマンスの向上を支援することができるとしている。

上記は主にユーザからのデータ収集であるが、開発者から開発に関するデータを収集する場合もある。開発者個人の作業時間や作業履歴のデータを自動的に収集するソフトウェアがいくつか公開されており、例えば Taskpit[11]では開発者が利用しているアプリケーションとその利用時間を自動的に計測する。

^{†1} 近畿大学大学院
Graduate School of Science and Engineering, Kindai University, Japan

^{†2} 近畿大学理工学部
Faculty of Science and Engineering, Kindai University, Japan

a) m.yukasa@gmail.com

b) tsunoda@info.kindai.ac.jp

- 🔄 アプリを再起動
- 📧 フィードバックを送信
- 🔇 端末が再起動するまでミュート

(a) クラッシュ時



(b) 送信フォーム

図 1 Android OS 上でのフィードバック送信画面

これらの収集されたデータは、ソフトウェアの品質改善や、ソフトウェア開発の効率向上に利用される。例えばクラッシュレポートのデータに関して分析した研究は非常に多い[1][2][10]。開発者の作業時間などのデータは、開発工数の見積もりへの利用や、ソフトウェアプロセスの改善(パーソナルソフトウェアプロセスなど)へ利用することが可能である。例えば、前出の Taskpit を実際のソフトウェア開発に適用し、開発者の不適切な振る舞い(管理者が開発者のようなタスクを行っている、電子メールの扱いが不適切など)を検出できた事例が紹介されている[11]。

上述のように、ソフトウェアに関するデータがユーザから収集され、これらを活用する研究が盛んになりつつある。ただし、データを収集するためには、ユーザの同意を必要とするため、すべてのユーザからデータを収集できているわけではない。企業に所属する開発者の場合、必ずしもデータ収集に関する同意は必要ではないが、できるだけ同意を得たほうが、開発者のモチベーションの低下を防ぐことができる。

そこで本稿では、ソフトウェアに関するデータのユーザからの収集を促進することを目指す。そのためのアプローチとして、ユーザに対してアンケートを実施し、データ収集に関する態度などを調査するとともに、データ収集に対するユーザの協力姿勢に影響する要因を分析する。

2. アンケートの設計

2.1 データ収集に対する協力姿勢

アンケートでは、データの収集にどの程度のユーザが応じるのかを調べた。データの収集は、プライバシー関連のデータの漏洩をユーザが連想しやすいため、すべてのユーザがデータ収集に同意するとは限らない。アンケートにおいて、以下の3種類のデータ収集に対する態度調査を行った。

- (1) クラッシュレポートへの協力姿勢
- (2) アプリケーションの操作ログへの協力姿勢
- (3) ソフトウェア開発の操作ログへの協力姿勢

(1)に関しては多くのユーザがクラッシュレポートの送信確認を見たことがあると考え、以下の4つの選択肢から実際の行動を回答してもらった。

- 気づいていなかった
- 特に気に留めていない
- 意図的に送信している
- 意図的に送信していない

なお調査時には、一部のアプリケーションでのみクラッシュレポートを送信しているか、全てのアプリケーションで送信しているかなどを質問したが、分析時にはそれらをまとめて集計した。また、選択肢の「気づいていなかった」は、そもそもクラッシュレポートの送信フォームが現れることを知っているかどうかを確かめている。

(2), (3) (特に(3))については、全てのユーザが実際にその場面に遭遇しているとは限らないため、想定する状況を説明し、取りうる行動を以下の4つの選択肢から回答してもらった。

- 協力したくない
- できれば協力したくない
- 周囲のユーザが協力するなら協力してもよい
- 積極的に協力したい

「周囲のユーザが協力するなら協力してもよい」については、予備調査の回答で見られたため、選択肢として加えた。

2.2 協力姿勢の理由

データの収集に協力的、または非協力的な理由についても調査した。理由については、前節の(1), (2), (3)ごとに多少異なると考えられるため、それぞれに対して選択肢を若干変更して準備した。選択肢についてはそれぞれの分析結果で示す。理由は複数選択可能とした。

2.3 データ収集元への信頼

ユーザがデータを送るということは、ユーザの立場からは、データの漏洩リスクが生じると考えることができる。もしユーザが、データ収集に対してリスクが低いとみなせば、データ収集に協力する可能性が高くなる。リスクが低いとユーザにみなしてもらう方法として、ユーザに対して収集されるデータや収集の仕組みを説明し、理解してもらうことがあげられる。

一方、リスクコミュニケーション(ステークホルダ間で化学物質などのリスクに関して、正確な情報を共有すること)においては、リスクの認識に対して、相手に対する信頼が影響すると指摘されている[6]。例えばデータ収集の場

表 1 クラッシュレポートへの協力姿勢

回答	割合
気づいていなかった	5%
特に気に留めていない	50%
意図的に送信している	20%
意図的に送信していない	25%

表 2 クラッシュレポートへの協力姿勢
 (信頼度が高い場合)

回答	割合
特に気に留めない	60%
意図的に送信している	25%
意図的に送信していない	15%

表 3 クラッシュレポートへの協力姿勢の理由

理由	割合
許可しても不利益にならない	30%
匿名性が保たれるかわからない	20%
誰もが許可していると思うため	15%
協力するメリットがない, もしくはわからない	15%
他の人が協力すれば十分である	10%
データ収集の効果を実感したことがある	5%
協力するメリットがある	5%
協力を求められた	5%
他の人の役に立ちたい	5%
知らされていないデータが収集されるかもしれない	5%
匿名性が保証されている	5%
データ収集の効果を実感できない	0%
信頼できるデータ収集元である	0%

合, データの匿名化方法はいくつか提案されているが [4][12], 匿名化方法と収集されるデータが明示されたとしても, そもそもデータの収集者が信頼できなければ (他のデータも収集される可能性を否定できなければ), ユーザが応じる可能性は低くなる。

ユーザの相手に対する信頼がどの程度影響しているかを明らかにするために, 2.1 節のアンケート時に, 想定するデータ収集者 (例えば Mozilla や Microsoft など) に対する信頼度を回答してもらった。文献 [6] と同様に 7 段階 (0: 全く信頼できない ~ 6: 非常に信頼できる) で回答してもらうこととした。さらに, 信頼度が高ければデータ収集に対する協力姿勢を変えるのかどうかについて調べた。そのために, データ収集者に対する信頼度を 5 または 6 と指定し, この場合の態度を 2.1 節で示した選択肢より回答してもらった。

これまで, ソフトウェア開発における信頼に着目した研究はいくつか存在する。具体的には, ソフトウェアを外部委託する場合 [7][8] や, オフショア開発する場合 [5] など, よく知らない相手と一緒にソフトウェア開発を行う場合の信頼の重要性について分析している。オンラインでチームを組む場合における信頼の重要性に着目した研究もある

[3]。ただし, データの収集元に対するユーザの信頼度が, データ収集の協力姿勢にどのような影響を与えるかを分析した研究は, 我々の知る限り存在しない。

2.4 調査対象者

アンケート調査の対象は, 大学で情報科学を専攻する学部 4 年生から修士課程 2 年生の 21 名 (2.1 節のアンケート (1) のみ 20 名) である。ソフトウェア工学の実験において, 学生を開発者の代替として被験者にしても結果に大きな差がないことが指摘されている [9]。そこで本稿でも, 比較的アンケート調査の実施が容易な学生を調査対象とすることとした。

なお, 情報科学を専攻していないユーザの場合, ソフトウェアに関する知識が相対的に少ないため, データ収集に対する認識が異なる可能性がある。そのようなユーザの割合は実際にも多いと考えられるため, これらのユーザについても調査対象とすることは今後の課題である。

3. 分析

3.1 クラッシュレポートへの協力姿勢

クラッシュレポートに対するユーザの姿勢を分析した。ここでは各自がプライベートで使用する端末において, クラッシュ時のデータを送信するタイプのクラッシュレポートに関するものとした。集計結果を表 1 に示す。クラッシュレポートに気づいていないという回答は 10% であり, 少なくとも多くのユーザはクラッシュレポートの存在には気づいているといえる。ただし, 特に気に留めていないユーザが半数の 50% であり, 存在を強く意識しているわけでない。意図的に送信しているユーザは 25%, 意図的に送信していないユーザは 20% であり, 協力的, 非協力的なユーザはどちらも多数ではなかった。

上記のアンケートでは, データの収集元に対する信頼度の平均値は 3.52 であり, 中程度であった。2.3 節で説明したように, 信頼度が高いと仮定した場合の協力姿勢について調査した。結果を表 2 に示す。意図的に送信しないという回答が減少し, 特に気に留めないユーザが若干増加したが, 協力的なユーザの割合は増加しなかった。このことから, データの収集元の信頼性が高いだけでは, クラッシュレポートのデータ収集に対する積極的な協力は得られないといえる。

ユーザの協力姿勢の理由を表 3 に示す。非協力的でない理由として最も多かったものは「クラッシュレポートの送信を許可しても自身に不利益になることがないから」であった。また, 「みんなが許可していると思うから」も比較的多かった。また, 非協力的な理由として多かったものは「匿名性が保たれるかわからないから」であった。「協力するメリットがない, もしくはわからないから」がその次に多かった。これらの結果より, ユーザにクラッシュレポートを送信するメリットが明確に伝わっていないといえる。

表 4 アプリケーションの操作ログへの協力姿勢

回答	割合	割合 (信頼度高)
協力したくない	38%	14%
できれば協力したくない	38%	48%
周囲のユーザが協力するなら協力してもよい	19%	33%
積極的に協力したい	5%	5%

表 6 ソフトウェア開発時の操作ログへの協力姿勢

回答	割合	割合 (信頼度高)
協力したくない	10%	5%
できれば協力したくない	62%	52%
周囲のユーザが協力するなら協力してもよい	19%	24%
積極的に協力したい	10%	19%

表 5 アプリケーションの操作ログへの協力姿勢の理由

理由	割合
匿名性が保たれるかわからない	52%
収集されるデータが不明確である	48%
知らされていないデータが収集されるかもしれない	48%
データ収集の効果が不明	43%
協力するメリットがない、もしくは不明	43%
他の人が協力すれば、それで十分である	29%
自分のデータを提供することで改善が期待できる	19%
改善するまでの時間が早くなるかもしれない	19%
信頼できる企業の提案なら協力したい	14%
記録することで報告の手間がなくなる	10%
自分のデータを1つのデータとして使ってほしい	5%
こだわりがなく、替えのきくアプリケーションである	0%
自分のデータは特殊なため、送信すると相手が混乱する	0%
自分が協力することで周りにも働きかけた	0%
依頼される前から自分も同じ意見を持っていた	0%
使い続けたい、もしくは替えのきかないアプリケーションである	0%

表 7 ソフトウェア開発時の操作ログへの協力姿勢の理由

理由	割合
データ収集の効果が不明	48%
知らされていないデータが収集されるかもしれない	29%
収集されるデータが不明確である	24%
匿名性が保たれるかわからない	24%
信頼できる企業や人の提案なら協力したい	24%
自分のデータを提供することで改善が期待できる	19%
自分の振る舞いを客観的に評価してほしい	19%
協力するメリットがない、もしくは不明	14%
他の人が協力すれば、それで十分である	14%
協力しないことで、周囲から批判的に見られるかもしれない	14%
改善に要する時間が短縮されるかもしれない	14%
周囲がどのような振る舞いをしているのか気になる	10%
自分が協力することで周りにも働きかけた	5%
自分のデータは特殊なため、送信すると相手が混乱する	0%
提案される前から自分も同じ意見を持っていた	0%
自分の振る舞いに自信がある	0%

3.2 アプリケーションの操作ログへの協力姿勢

2.1 節で説明したように、アンケートでは「アプリケーションを運営する企業から、ユーザビリティ向上のため、今後アプリを使用する際に操作やログなどを記録させてほしいと依頼があった」という前提で回答してもらった。集計結果を表 4 に示す。アプリケーションの操作ログの場合、「協力したくない」と「できれば協力したくない」の回答が多くを占め、76%となった。データの収集元に対する信頼度の平均値は 3.19 であった。

次に、データの収集元への信頼度が高いとした場合の回答を表 4 の「信頼度高」の列に示す。3.1 節の結果と異なり、信頼度が高い場合にユーザの姿勢の変化が見られた。「協力したくない」が 14%に低下するとともに、「周囲のユーザが協力するなら協力してもよい」と「積極的に協力したい」の回答の合計が 38%となった。このことより、アプリケーションの操作ログの収集に関しては、データの収集元に対する信頼が最も重要であるといえる。

協力姿勢の理由を表 5 に示す。収集対象が操作ログであるため、データの匿名性などに関する不安回答が多かった。また、データ収集に協力するメリットが不明であるという回答も多かった。3.1 節でもメリットが不明であるという回答が多かったが、操作ログはプライバシー漏洩のリスクが高いと連想されやすいデータのため、メリットが不明であることが非協力的な態度につながった可能性がある。

3.3 ソフトウェア開発時の操作ログへの協力姿勢

ここでは、「ソフトウェア開発をチームで行うとし、業務改善（プロセス改善）のために、開発時のキーストロークやマウス操作などを記録することが提案されている」とした。この場合のデータ収集への協力姿勢を表 6 に示す。「協力したくない」は多くなかったが、「できれば協力したくない」という回答が多く、62%となった。データの収集元に対する信頼度の平均値は 3.33 であった。

データの収集元への信頼度が高い場合の協力姿勢を表 6 の「信頼度高」の列に示す。高い信頼度を設定しなかった場合（左側の列）、データの収集に対して協力姿勢を示し

た回答（「周囲のユーザが協力するなら協力してもよい」と「積極的に協力したい」の回答の合計）は29%であったが、43%に上昇した。このことから、ソフトウェア開発時の操作ログ収集への協力姿勢についても、データ収集元への信頼が影響しているといえる。

協力姿勢に関する理由を表7に示す。3.2節の場合と信頼度の平均値に大きな差はなかったが、データの収集内容に関する懸念を示した回答は少なかった。「協力したくない」という回答が少なかったのは、このことが原因である可能性がある。ただし、データ収集の効果が明らかでないという回答の割合は多く、このことが「できれば協力したくない」という回答の多さに繋がった可能性がある。3.1節の場合と同様に、この場合もデータ収集の効果を相手に説明する必要があるといえる。

3.4 考察

クラッシュレポートのデータ収集に関してのみであるが、自身がより協力したくなる、もしくは、非協力的な人に対して、データ収集元はどのような取り組みをしたら良いと思うかについて自由記述で答えてもらった。回答された内容には、主に以下の2つが多くみられた。

- (1) 明確な記述が欲しい
- (2) 自分に利益があるとよい

(1)については、例えばアップデートの際にどのようなクラッシュレポートが参考にされたのか、協力した場合の利益を明確にするなどの記述が多く見られた。これはこれまでの分析結果とも一致するものである。(2)については、ポイントの付与や、協力者のみに一部の機能やサービスの解放をするなどがあげられた。その他の回答として、ユーザビリティ改善（クラッシュレポートの送信タイミングが適切でない）をあげている回答もみられた。

4. おわりに

本研究では、ソフトウェアに関するデータのユーザからの収集を促進することを目的とし、ユーザに対してデータ収集に対する態度のアンケートを実施し、ユーザの協力姿勢に影響する要因を分析した。収集するデータとしては、クラッシュレポート、アプリケーションの操作ログ、ソフトウェア開発の操作ログの3つを想定して、それぞれに対する態度を回答してもらった。また、データ収集に協力的（または非協力的）な理由と、データの収集元に対する信頼の影響について分析した。分析の結果、以下の傾向が見られた。

- クラッシュレポートに関しては、ユーザは存在を強く意識していなかった。データ収集元への信頼度が高い場合、データを送信しないユーザは減少したが、積極的に送信するユーザは増えなかった。

- アプリケーションの操作ログについては、信頼度が高くない場合、データ収集に協力的でない回答の割合が多かった。信頼度が高い場合、データ収集への協力姿勢に改善が見られた。
- ソフトウェア開発の操作ログについても、データ収集元への信頼度が高いと、協力姿勢が改善する傾向が見られた。
- ユーザのデータ収集への態度の理由として、収集されるデータや収集に協力した場合の効果が明確でないという理由が多かった。これらを明確にすることにより、協力姿勢が改善する可能性がある。

本研究の主要な貢献のひとつは、「データ収集元に対する信頼」に着目して分析を行ったことと、収集するデータの種類によっては、信頼の影響の度合いが異なることを示したことである。今後の予定は、実際に収集されるデータや収集されるメリットを明示することにより、ユーザのデータ収集への協力姿勢が変化するかを確かめることである。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金（基盤C:課題番号16K00113, 基盤A:課題番号17H00731）による助成を受けた。

参考文献

- [1] Campbell, J., Santos, E., and Hindle, A.: The unreasonable effectiveness of traditional information retrieval in crash report deduplication, In Proc. of International Conference on Mining Software Repositories (MSR), pp.269-280 (2016).
- [2] Castelluccio, M., Sansone, C., Verdoliva, L., and Poggi, G.: Automatically analyzing groups of crashes for finding correlations, In Proc. of Joint Meeting on Foundations of Software Engineering (ESEC/FSE), pp.717-726 (2017).
- [3] Coppola, N., Hiltz, S., and Rotter, N.: Building trust in virtual teams, In Proc. of Professional Communication Conference (IPCC), pp.353-366 (2001).
- [4] Microsoft Corporation: Collecting anonymous and traceable telemetry, US patent, US8607305 B2 (2008).
- [5] Moe, N., and Šmite, D.: Understanding a lack of trust in Global Software Teams: a multiple-case study, Software Process: Improvement and Practice, vol.13, no.3, pp.217-231 (2008).
- [6] 中谷内一也: リスク管理への信頼と不安との関係—リスク間分散に着目して—, 心理学研究, vol.82, no.5, pp.467-472 (2011).
- [7] Nguyen, P., Babar M., and Verner, J.: Critical factors in establishing and maintaining trust in software outsourcing relationships, In Proc. of international conference on Software engineering (ICSE), pp.624-627 (2006).
- [8] Oza, N., Hall, T., Rainer, A., and Grey, S.: Trust in software outsourcing relationships: An empirical investigation of Indian software companies, Information and Software Technology, vol.48, no.5, pp.345-354 (2006).
- [9] Salman, I., Misirli, A., and Juristo, N.: Are students representatives of professionals in software engineering experiments? In Proc. of International Conference on Software Engineering (ICSE), pp.666-676 (2015).

- [10] Seo H., and Kim, S.: Predicting recurring crash stacks, In Proc. of International Conference on Automated Software Engineering (ASE), pp. 180-189, (2012).
- [11] Suthipornopas, P., Leelaprute, P., Monden, A., Uwano, H., Kamei, Y., Ubayashi, N., Araki, K., Yamada, K., and Matsumoto, K.: Industry Application of Software Development Task Measurement System: TaskPit, IEICE Transactions on Information and Systems, vol.E100-D, no.3, pp.462-472 (2017).
- [12] Yang, Z., Zhong, S., and Wright, R.: Anonymity-preserving data collection, In Proc. of SIGKDD international conference on Knowledge discovery in data mining (KDD), pp.334-343 (2005).