

オンラインアンケートの回答信頼性に影響する指標の調査検討

- アンケート形式とスマートフォン操作状況の観点から -

Considerations on affectors of reliability for answer to online questionnaires

- From point of style of questionnaire and context of smart phone manipulation -

後上 正樹[†] 谷 優里[†] 松田 裕貴[†] 荒川 豊[‡] 安本 慶一[†]
Masaki Gogami Yuri Tani Yuki Matsuda Yutaka Arakawa Keiichi Yasumoto

1. はじめに

アンケート調査では回答者が必ずしも正確に回答するとは限らない。適切な認知コストを払わずに回答された結果から導かれる推論は無意味なものになってしまうため、それを検出・排除できないことはアンケート調査における致命的な問題と言える。近年、クラウドソーシングによるオンラインアンケート調査が社会科学関連の研究にも用いられている。これは、インターネットを介して全世界の参加希望者を対象としてアンケートの回答を依頼することができる仕組みである。これを用いることで、一度に大量のデータを取得できる上に、比較的容易にデータの代表性を保障できる等の利点が期待される。

Simon ら [1] は、人間の認知的資源には限りがあることによって生じる、アンケート調査において回答要求に対する努力を最小化しようとする傾向を努力の最小限化 (satisfice) と定義した。また、Krosnick ら [2] は satisfice の分類として、選択肢を十分に検討せず回答する場合を弱い satisfice、全く認知コストを払わずあてずっぽうに回答するような場合を強い satisfice と分類した。三浦ら [3] は、IMC (Instructional Manipulation Check) や DQS (Directed Question Scale) という、いわばひっかけ問題のようなものを設問文や選択肢に混入したアンケートを作成し、回答結果から強い satisfice の検出を試みた。しかし、この方法ではひっかけ問題を配置した箇所以外での satisfice が検出できない。また、設問数が増えるため、回答者の負担が増加するという欠点も考えられる。

近年、オンラインアンケートの回答に用いられるデバイスとして、スマートフォンが PC に取って代わってきている。これを受けて、Roger ら [4] は、PC およびタブレットとスマートフォンでアンケート結果の質にどのような違いがあるのかを調査し、結果の信頼性についてどのデバイスについても特に差はないと結論づけている。したがって、スマートフォンは他のデバイスと比べてオンラインアンケートを回答するデバイスとして不適切でないと言える。また、その高いポータビリティにより今



払われる認知コスト：少
→ satisfice：あり
→ 結果の信頼性：低

払われる認知コスト：多
→ satisfice：なし
→ 結果の信頼性：高

- (a) 払われる認知コストが少なく、結果の信頼性が低い例
(b) 払われる認知コストが多く、結果の信頼性が高い例

図 1: スマートフォンの操作状況がアンケート結果の信頼性に影響を与える例

後さらにオンラインアンケートの回答に用いられると予測される。

そこで本研究では、図 1 のようなオンラインアンケートをスマートフォンで回答する際に払われる認知コストが少ないことによって発現する satisfice を検出することを目標とする。satisfice の検出ができれば、信頼性が低い回答結果を抽出して排除することができ、アンケート結果を用いた推論の精度向上に貢献することができる。

その目標に際して本稿では、スマートフォンによってアンケートを回答する際の satisfice に影響を与える要因と satisfice を検出するための指標について検討する。また、その指標を定量的に記録する実験を設計したので報告する。

本稿の構成は次の通りである。2 章では関連研究について述べる。3 章で satisfice に影響を与える要因について検討し、4 章で satisfice を検出するための指標について述べる。5 章では、4 章で検討した指標を取得するアンケートシステムの構想について述べる。最後に、6 章で本稿のまとめと今後の展望を述べる。

[†] 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

[‡] 九州大学, Kyushu University

2. 関連研究

Rogerら [4] は、PCおよびタブレットとスマートフォンでアンケート結果の質にどのような変化があるのかを調査した。この際、評価基準としたのは回答時間や未回答率および題意に沿わずある一定の規則に沿って回答する行為を指すストレートライニングである。結果として、スマートフォンはPCおよびタブレットに比べて回答時間が長い傾向が観察された。しかし、結果の信頼性についてはどのデバイスについても特に差はないと結論づけている。

Janら [5] は、選択肢がどの設問でも固有で賛成と反対の要素で構成されるA/D (Agree/Disagree) 形式と選択肢が設問ごとに対応付けて設定されているIS (Item-Specific) 形式のアンケートを比較する実験を実施した。アイトラッキングによって回答時の視線を記録したところ、IS形式の方が注視回数が多く、注視時間が長く、再注視回数も多かったと報告した。

加藤ら [6] は、スマートフォンにおけるwebブラウジング中の平均スワイプ速度及びユーザの注視時間を取得することで、閲覧中の記事への興味度合いを推定した。結果として、これらの特徴量と興味度合いを尋ねるアンケート結果に強い相関が確認された。これより、Janら [5] の結果から、A/D形式よりもIS形式の方がアンケートに対する興味度合いが高くなる可能性がある。

平部ら [7] は、スマートフォンに内蔵されている種々のセンサに加えて、タッチ操作をセンサとして活用するシステムをAndroid端末において構築した。8種類のタッチ操作を識別する実験において、識別精度は85%であったと報告している。また、例えば片手で持つ場合や両手で持つ場合等の8種類のスマートフォンの持ち方が96%の精度で推定できたことを報告している。

Marioら [8] は、30種類の操作特徴セットを用いて、上下左右の基本的なスワイプによって個人の操作特徴を分類及び識別することで新しいバイオメトリック認証の可能性を検討した。実験では、個人の操作を登録してから1週間後の認証テストにおいて平均誤り率が4%未満であったと報告している。平部らとMarioらの技術を用いれば、アンケートの回答者ごとに丁寧に回答した時の操作を記録し、それ以降の回答時の操作と比較することでsatisficeを検出することができる可能性が考えられる。

3. satisfice に影響を与える要因の検討

satisfice は正確に回答しようとする意志の欠如によって生じる努力の最小限化である。この欠如の要因としてスマートフォンの操作状況とアンケート形式の観点から考えられる事項を表1に示し、各事項について以下で詳細を述べる。

表 1: satisfice に影響を与える要因の例

要因カテゴリ	要因
スマートフォン操作状況	回答時間帯
	ながら操作
アンケート形式	設問の提示方法
	選択肢の提示方法

回答時間帯 例えば、次の予定があり十分に時間がない朝の時間帯であれば、アンケートに割ける時間が短いことからsatisficeが発現し易い傾向が見られる可能性がある。一方で、次の予定に追われることなく落ち着いた状態でアンケートに取り組むことができる夜には、satisficeが発現しにくい傾向がみられる可能性が考えられる。

ながら操作 回答中の回答者の状態は様々であり、それがsatisficeに影響する可能性が考えられる。例えば、食事中に回答する場合は食べ物とアンケートの両方に意識を向けることになるため、アンケートを答える行為だけを行う場合よりもsatisficeは発現しやすい可能性があると考えられる。このように、オンラインアンケートにおいては何か他のことをしながら回答するケースは大いにあり得るため、ながら操作を考慮しなければならないと考える。そこで、平部ら [7] とMarioら [8] の結果を応用すれば、スマートフォンの持ち方とタッチスクリーン操作から個人ごとにながら操作を大まかに推定できる可能性が考えられる。

設問の提示方法 設問の提示方法についてもsatisficeに影響する要因となる可能性がある。例として、図2のような2種類の提示方法を考える。左側の提示方法は1ページの中に複数の設問が表示される場合である。回答者は画面を縦にスクロールすることによって次の設問に遷移する。一方で、右側の提示方法は1ページにつき1つの設問が表示される場合である。次の設問へ遷移するには、画面右下にある「次へ」ボタンをタップし、次のページを読み込む必要がある。なお、ブラウザの「戻る」ボタンによって前問に戻ることも可能である。これら2種類の提示方法の違いは、回答の変更行動に影響を与える可能性が考えられる。例えば、3問目を回答中に2問目の回答を変更したくなった場合、左側のアンケートでは上にスクロールするという比較的瞬時性が高い操作で2問目の回答を変更することができる。しかし、右側のアンケートでは、2問目の設問や選択肢は一切見えておらず、ブラウザの「戻る」ボタンをタップして再度2問目の回答ページを読み込むため、左側のアンケートより



(a) 1 ページに複数問表示される形式 (b) 1 ページに 1 問だけ表示される形式

図 2: 異なる設問の提示方法の例

も回答の変更にかかる。そう長くはないこの時間が大きな心理的障壁となり、回答を変更すべきという想起を実行に移すことを辞めてしまう可能性が考えられる。

選択肢の提示方法 設問の提示方法と同様に選択肢の提示方法についても *satisfice* への影響が考えられる。例として図 3 に示すような 2 種類の提示方法が挙げられる。左側の図は図 2 の右側の図と同一のアンケートであり、選択肢を固定する A/D 形式である。一方で右側の例は設問文も選択肢も設問ごとに変化させる IS 形式である。Jan らの報告より、IS 形式である右側の提示方法の方が A/D 形式である左側に比べて *satisfice* の発現が抑制される可能性が考えられる。

4. *satisfice* を検出するための指標の検討

satisfice を検出するための指標として考えられる事項を表 2 に示し、特に着目する事項について以下で詳細に述べる。また、5 章で説明する評価実験に用いるアンケートシステムである LimeSurvey [9] について、表中では LS と略記する。

アンケートの回答時間は設問文および選択肢の文章量と難易度および回答者の傾注度合いに依存すると考えられる。したがって、「アンケート全体の回答時間」と「設問単位の回答時間」が、それぞれユーザー単位および設問単位の *satisfice* と関連している可能性が考えられる。ここで、文章量と難易度は回答者に依存しない要素であ



(a) 選択肢を固定する形式 (b) 選択肢を設問ごとに設定する形式

図 3: 異なる選択肢の提示方法の例

表 2: *satisfice* が検出できる可能性を持つ指標の例

指標カテゴリ	指標	取得システム
回答方法	アンケート全体の回答時間	LS 標準機能
	設問単位の回答時間	LS プラグイン
	回答の変更	LS プラグイン
	スクロール速度	LS プラグイン
	タップ回数	LS プラグイン
回答環境	配信から回答までの時間	LS 標準機能
	回答時間帯	LS 標準機能
	ながら操作	アプリケーション

ると仮定する。このとき、Jan ら [5] と加藤ら [6] の結果から、設問ごとに文章量と難易度に依存する回答時間の閾値を定めることによって、それよりも明らかに速い回答について、回答者の興味度合いが低いとみなすことができる。ゆえにその回答は、適切な認知コストが払われず、*satisfice* が発現したと解釈できる可能性が考えられる。

「回答の変更」は、少なくとも意図して雑な回答をしようとした場合には発生しないと考えられる。したがって、少なくとも回答の変更が見られた回答は *satisfice* が発現していないとみなすことができると考える。ただし、変更されなかった回答に払われた認知コストが必ずしも低いとは言えないため、この指標のみでの的確に *satisfice* の検出が実現できるわけではない。正確に回答した結果も含めて、選択肢を変更しなかった回答を除外すれば高純度な結果を得ることはできるが、データ数の大幅な減少が予想されるため、選択肢が変更される割合についても分析する。

「配信から回答までの時間」は回答者のアンケートに

対する積極性を反映する可能性が考えられる。

「回答時間帯」と「ながら操作」については3章で述べたとおりである。

5. 評価実験のためのシステム構想

3章で述べた要因によって発現すると考えられる *satisfice* を4章で検討した指標によって検出可能かどうかを評価する実験の構想について以下に述べる。

5.1 システム構成の検討

システムの要件としてアンケートの実施と様々な指標のセンシングの2点が挙げられる。しかし、LimeSurveyの標準機能では回答操作ログを取得することができない。そこで、要件を満たすシステム構成を検討する上で、機能をLimeSurvey上に実装する場合と、クライアント側のスマートフォンアプリケーションとして実装する場合のそれぞれについて利点と欠点を以下に述べる。

LimeSurvey上で構成する利点として、多様なアンケート設定を実現する機能を持つアンケートシステムを一から実装する必要がない点が挙げられる。また、オープンソースのシステムであるため、独自の追加機能を開発・追加することができる。これにより、回答結果に紐づいた形で回答操作ログを取得することができる。一方で、スマートフォンに内蔵されているセンサのデータを取得することができないという欠点がある。また、クライアント側のアプリケーション上で実装する利点として、内蔵のセンサデータを取得できる点が挙げられる。しかし、アンケートシステムを一から実装する必要がある。

そこで、本研究ではサーバとクライアント両方の利点を備えたハイブリッドなシステムを実装する。

5.2 ハイブリッドな回答操作センシングシステム

本研究で実装するハイブリッドな回答操作センシングシステムを図4に示す。このシステムは、LimeSurveyをベースとして、それを包含するアプリケーションで構成される。LimeSurveyでは、アンケートを配信し、回答操作のセンシングを行う。アプリケーションでは、配信されるアンケートを実施し、回答時のスマートフォンの持ち方や揺れ等を検出するために各種内蔵センサのデータを記録する。また、回答時のタッチスクリーン操作を記録するために、LimeSurveyのプラグインを開発する予定である。アプリケーションおよびLimeSurveyで取得するデータは全てデータベースに格納する設計を計画している。

5.2.1 LimeSurveyでのセンシング

表2の取得システムの欄にLS標準機能と記載されている「アンケート全体の回答時間」、「配信から回答までの時間」、「回答時間帯」の3項目については、LimeSurvey

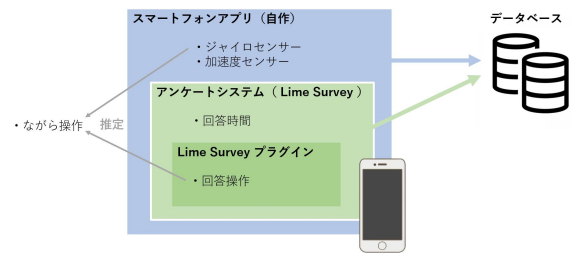


図4: 評価実験のシステムおよび取得するデータ

の標準機能で取得可能である。しかし、LSプラグインと記載されている項目についてはLimeSurveyの標準機能では取得できない。そこで、それらの項目のデータを取得する機能をプラグインとして開発する。タッチスクリーンの操作を取得する方法として、JavaScriptのタッチイベントを用いる。これにより、指がスマートフォンの画面に触れていることを検知し、そのときの時刻と画面上の座標を取得することができる。また、特定のDOM (Document Object Model) 要素が現在画面に表示されているかどうか及び画面内の座標を1ピクセル単位で取得することができるAPIであるInteraction Observer API (以下、IO-APIと略記する)も併せて用いる。まず、現在どの画面が表示されているのかという情報をIO-APIによって取得する。それに加えて選択肢をタップしたことをタッチイベントによって取得することで「設問単位の回答時間」を検出する。同様の手法で、一度回答した設問の他の選択肢をタップしたことをタッチイベントによって取得することで「回答の変更」を検出する。「スクロール速度」については、IO-APIによってスクロールによる表示範囲の差分とそれに要した時間から算出する。「タップ回数」についてはタッチイベントを単純にカウントすることで取得可能である。

5.2.2 アプリケーションでのセンシング

LimeSurveyのプラグインで取得する回答操作ログを用いて、平部ら[7]とMarioら[8]の技術により、スマートフォンの持ち方とタッチスクリーン操作から個人ごとにながら操作を大まかに推定できる可能性が考えられる。また、アプリケーションで取得するジャイロセンサーと加速度センサーのデータにより、回答時のスマートフォンの揺れを検出する。これによって、ながら操作の推定精度を向上することができる可能性がある。

5.3 アンケート設計

アンケートの設計として、設問項目と回答方法それぞれについて検討する。

設問項目について、記憶を辿るタイプの設問を検討している。回答に必要な記憶情報の時間的深さを調整することで、設問の難易度を調節することができる。考える。

また、膨大な設問量をこなすタスクであるという表記を途中の段階で挿入することで、それ以降の回答における satisfice を調整できる可能性が感られる。さらに、前問に戻ることが可能なアンケートにおいて、条件分岐によって設問数が大幅に増加した場合に satisfice が発現し、分岐点の回答を変更する可能性が考えられる。提示方法については、3章で述べたように、1ページに1問表示する場合と複数問表示する場合を組み込むことを検討している。

回答方法については、3章で述べたような選択肢を固定する場合と設問ごとに設定する場合の比較要素を組み込む。ただし、全ての選択肢において数値等の定量的な尺度ではなくリッカート尺度を用いることとする。

5.4 実験手法

本稿で注目している satisfice はその性質上、実験室で行われる一般的な実験では検出することができない可能性がある。実験室においてアンケートを行うよう参加者に教示すると、集中して回答を行ってしまうためである。ゆえに、本評価実験は satisfice が発現する状況下で行う必要がある。そこで、日常生活の中でアンケートの回答依頼通知が参加者に届くというシステムを提案する。また、グラウンドトゥルースの収集方法としては、アンケート回答者に設問文と選択肢、および回答結果を提示し、その回答に至った理由等のデプスインタビューを回答者自身で行うシステムの運用を検討している。

6. まとめ

本研究では、オンラインアンケートの回答に用いる端末として今後さらに広く採用されるであろうスマートフォンによる回答結果の信頼性について、アンケート形式と回答操作状況の観点から検討した。今後は提案したプラグインおよびアプリケーションを実装し、操作ログが結果の信頼性に対して有用なデータであるかどうかを評価する実験を行う。

また、本研究では、アンケートの回答結果の信頼性が損なわれる原因として satisfice を取り上げたが、他にも反応バイアスというものがある。これは、正確に回答する意思の有無に関わらず、ヒトの心理的性質に起因する統計誤差である。本研究で取得する予定の操作ログを用いて反応バイアスの検出を実現し、さらなるデータの信頼性向上に貢献することが期待される。

謝辞

本研究の一部は、JST さきがけの支援のもと実施されている。

参考文献

- [1] Herbert A. Simon. Rational choice and the structure of the environment. 1956.
- [2] Jon A. Krosnick. Response strategies for coping with the cognitive demands of attitude measure in surveys. pp. 213–236, 1991.
- [3] 三浦麻子, 小林哲郎. オンライン調査における努力の最小限化が回答行動に及ぼす影響. 行動計量学, Vol. 45, No. 1, pp. 1–11, 2018.
- [4] Roger Tourangeau, Hanyu Sun, Ting Yan, Aaron Maitland, Gonzalo Rivero, and Douglas Williams. Web surveys by smartphones and tablets: Effects on data quality. *Social Science Computer Review*, Vol. 36, No. 5, pp. 542–556, 2018.
- [5] Jan Karem Hhne and Timo Lenzner. New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions. *Journal of Survey Statistics and Methodology*, Vol. 6, No. 3, pp. 401–417, 10 2017.
- [6] 加藤勇太, 岩本健嗣, 松本三千人. タッチ操作ログを用いた web コンテンツ閲覧時における興味度合い推定の研究. 情報処理学会論文誌, Vol. 59, No. 2, pp. 508–518, 2018.
- [7] Yuko Hirabe, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. Touchanalyzer: A system for analyzing user's touch behavior on a smartphone. *International Journal of Computer Science and Mobile Computing*, Vol. 7, pp. 25–38, 2018.
- [8] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 1, pp. 136–148, Jan 2013.
- [9] Limesurvey: the online survey tool - open source surveys. <https://www.limesurvey.org/>. (Accessed on 07/12/2019).