

回答の内的一貫性に基づくWEB調査における不良回答の自動識別に関する研究

メタデータ	言語: ja 出版者: 静岡大学 公開日: 2014-02-04 キーワード (Ja): キーワード (En): 作成者: 山田, 文康 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/7526

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 3 月 31 日現在

機関番号：13801

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500255

研究課題名（和文） 回答の内的一貫性に基づくWEB調査における不良回答の自動識別に関する研究

研究課題名（英文） Method of distinguishing defective responses from web survey data on the basis of internal consistency of responses

研究代表者

山田 文康 (YAMADA FUMIYASU)

静岡大学・情報学部・教授

研究者番号：40158217

研究成果の概要（和文）：

一般に調査結果の一部には、質問文をよく読まないで回答する「不良回答」が含まれているとされ、それは特に Web 調査で問題視されている。本研究では SD 法データを対象とし、回答の内的一貫性に基いて不良回答を識別する手法を開発した。取得したデータに適用した結果、回答時間が短いなど不良性の特徴を持つ回答者を識別でき、本手法が適切に不良回答候補を選別し得ていることを確認できた。また、一般的な 2 相データへの拡張する方法についても検討を加えた。

研究成果の概要（英文）：

It is often said that replies to a questionnaire may include defective responses that are given by respondents who do not read questions, and that their ratio is higher especially in a web survey. In this paper, we developed a new method of distinguishing such kind of defective replies from a questionnaire data on the basis of internal consistency of responses. Applying our method to several semantic differential (S.D.) data, we could appropriately distinguish several respondents considered to be defective. And we also discussed about a method which can extend the coverage of data type to a general two mode questionnaire data.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
22年度	900,000	270,000	1,170,000
23年度	800,000	240,000	1,040,000
24年度	700,000	210,000	910,000
総計	2,400,000	720,000	3,120,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：Web 調査、不良回答、回答の内的一貫性、調査の信頼性と妥当性、Satisfice な回答

1. 研究開始当初の背景

近年、インターネットの普及や既存の調査法（訪問調査、郵送調査等）における調査環境の悪化等により、Web 調査が急速に普及してきている。しかし、Web 調査にはそもそも母集団が明確には特定出来ないという大きな問題がある。母集団が特定できないことは、標本調査が機能しないこと、従って、一般的な社会調査のためには利用できないことを意味する

もう一つ重要な問題は、Web 調査の回答結果の信頼性・妥当性の問題である。現在、大規模な Web 調査は民間の調査会社によって実施されているが、その場合、各社で募集したモニターが調査対象であり、回答に対しては謝礼（ポイント制）が支払われる。このようなモニター制は、モニターの属性を必ずしも正確に把握できないこと、謝礼目当ての回答が増える可能性があること等のため、回答の質の低下が懸念されている。さらに、Web 調査票には質問数が極めて多いもの、選択肢が非常に多い質問等も散見されている。これらの問題点は、Web 調査で得られた回答には、不正回答や虚偽回答に加え、質問文さえ読まない「不良回答」が無視できない比率で混入している可能性を示唆する。実際、調査会社の実施している Web 調査票では、「不良回答」を識別することのみを目的とした質問項目（回答するカテゴリを指定）を調査票に含めるケースも多くなっており、問題の深刻さを伺わせる。

回答の信頼性評価については、以前より心理学の分野で研究が進められており、特に性格検査ではライ・スケールがこの目的のために用いられている。しかしライ・スケールは、意図的に自分を良く見せかけようとする等の虚構反応を検出するために用意された尺

度であり、Web 調査で問題となる「不良回答」には適さない。元々、本研究で問題とする「不良回答」を識別する方法についてはほとんど研究がなされておらず、その対処方法としては、同一内容の項目を複数調査票に潜ませ、その回答の食い違いを調べる程度である。

2. 研究の目的

本研究で問題とする「不良回答」者は、質問文さえ読まないで回答する者であり、このような場合、回答は質問文の内容とは無関係になされる。そのため、本来、質問間に存在するはずの関連も生じず、結果として「不良回答」者における質問間の相関は、一般の回答者のそれとは大きく異なることが予想される。このような観点から、研究代表者を中心とし、設問数が多く、同じ形式の問い掛けが繰り返され、従って不良回答の生じやすい SD 法を取り上げ、データ解析を行った。その結果、不良回答が疑われる回答者においては、それ以外の回答者と比較して質問項目間の相関係数が大きく低下する事を見だし、さらに大学の講義に対する学生の評価データの解析から、無回答を含むデータの場合にも評価項目間の相関係数が低下することを見いだした。

このように、不良回答の場合には、質問内容に基づく内的な一貫性が損なわれ、それが項目間の関連を失わせる。従って、回答データの質問間の関連に基づけば、不良回答をある程度高い精度で識別することができるはずである。本研究は、SD データの解析から始まった「不良回答」識別研究を発展させ、その手順を統計的に厳密な方法として確立させるとともに、それを一般的な調査項目に拡張していくことを目的とする。

3. 研究の方法

Web 調査環境を模した実験システムを開発・利用し、多様な状況の下でのデータを複数取得する。これらのデータの解析に基づき、これまでの研究で得られた SD データにおける「不良回答」の識別方法を洗練するとともに、それを一般的な調査項目にも適用可能な方法へと拡張していく。さらに、ここで開発した方法を用いて、不良回答に関連する以下の特性についても検討する。

◆質問項目数と不良回答率との関係（項目数はどの程度までに抑えるべきか？）

◆質問形式と不良回答率との関係（不良回答が生じやすいのはどのような形式の質問か？）

◆不良回答の回答パターンの類型化（不良回答はどのようなパターンで現れるか？）

◆不良回答の分析結果への影響（不良回答の解析結果に与える影響はどのようなものか？）

4. 研究成果

(1) SD データを用いた不良回答識別の標準的な手順の開発

標準的な手順 は以下の通りである。

- ① 回答者ごとの質問（形容詞対）間相関の算出（回答者の評価視点の抽出）
- ② 回答者間の相関の算出（回答者間の評価視点の類似性の算出）
- ③ 回答者の評価視点の共有性の算出

回答者ごとに、回答者間相関係数が設定した基準値 α 以上の値を示した回答者数 S を算出する。 S が大きい回答者は他の多くの回答者と類似した意味空間を持っており、その意味で形容詞対の意味に基づいた回答（正常回答）を行っていると思える。逆に、 S が小さい回答者は他の多くの回答者とは異なった意味空間に準拠することになり、不良回答者の可能性が示唆される。

以上の手順を適用して得られた視点共有者数 S の分布例を図 1 に示す。データは、企業 15 社に対し、10 問の形容詞対を用いて取得した企業イメージであり、基準値 $\alpha = 0.5$ として分析した結果である。図によれば、分布は双峰性を示しており、視点共有者の多い「正常回答群」と極端に少ない「不良回答者候補群」に識別されていることが分かる。このような双峰性の分布は、適用した全てのデータにおいて得られた。

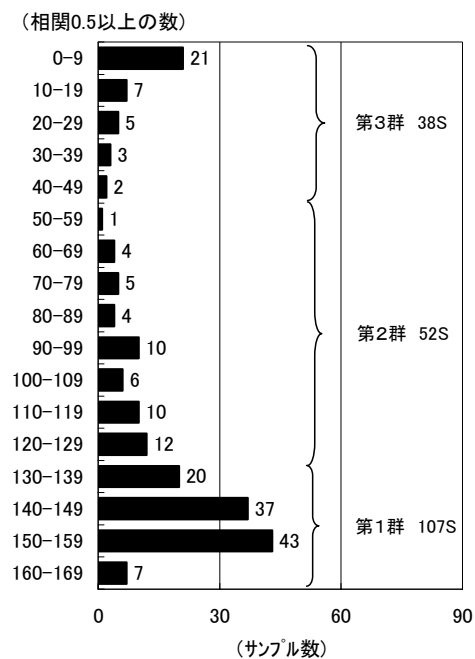


図 1. 視点共有者数の分布

(2) 異なった調査環境で取得したデータに基づく不良回答識別手順の評価

回答結果の質に影響を与える以下の 4 つの異なった環境で SD データを取得した。

- ① 管理された環境：不良回答の発生しにくい状況。
- ② 通常の Web 環境
- ③ 長大調査票を用いた調査：SD 調査に入る前に大量な項目群への回答を求めた
- ④ 不良回答データ：「調査項目を見ず、デタラメに回答」するように求めた。

この 4 データに対して不良回答識別手順を適用した結果、①～④の順で不良回答候補が

増えることが確認され、特に①については殆どが正常回答者、④は殆どが不良回答者として識別され、手法の妥当性が確認された。

(3) 調査票の形式の違いが与える影響

SD法の調査票（スポーツのイメージ）を用い、調査項目を回答者ごと無作為に構成・配置することが回答結果にどのような影響を与えるかを、固定した調査票で得たデータとの比較を通して実証的に明らかにした。調査項目の配置については、評価項目の提示順序、形容詞対の提示順序、さらには形容詞の尺度の両端への配置に関しても、回答者ごとランダムに設定した。

調査項目を無作為に配置した場合、それだけ回答者の負担が大きくなることが考えられる。分析の結果、その影響は回答に以下のような影響を与えていることが明らかとなった。

①回答時間が長くなる

②評価の視点が曖昧となる

- ◆中央のカテゴリに回答が集まる
- ◆形容詞対間相関がほぼ一定の値をとり、形容詞対による評価の差異が小さい
- ◆回答は「好意-非好意」の大枠での評価に留まり、詳細なイメージにまで引き出せていない。

調査票に対する回答は、回答者にとって自分の考えを整理し、再構成する作業でもあるが、本研究で行った形容詞対の並びと配置の無作為化は、回答者が評価の視点を構成していく過程を妨げた可能性がある。無作為化によって系列位置効果を排除することも重要であるが、同時に回答者がスムーズに自身の考えを整理していく流れを整えることも、項目の配置には重要な視点と考える。

(4) 一般的な質問項目への拡張

SDデータに対する不良回答識別法を、より

一般的な2相データに拡張する方法について検討を加えた。識別は、回答パターンの特徴を手掛かりとする手順と、回答の一貫性に基づく手順の2つから構成し、前者では回答所要時間、回答移動距離、中央カテゴリ集中傾向の3つの指標を用いた。また、後者については、相関の高い項目対群を抽出し、その項目対群に対する回答が主要な回答傾向とどの程度一致しているか、つまり、回答が相関係数に示される方向でなされているか否かを基準とした。一致度指標としては、一致の強さも考慮する指標（方法1）とその方向の一致性のみを評価する指標（方法2）の2つを作成した。適用データは72項目からなる性格検査で、回答不良が生じにくい環境（研究室）と生じる可能性のある環境（講義室）の2つで調査した。得られた結果の概要は以下の通りである。

①講義室環境で得たデータの一部には、回答パターンの特性として「不良性の高い」回答が含まれていることが確認できた。しかし、回答の一貫性、特に方法2で評価した場合には、研究室と講義室でほとんど差が見られないことも明らかとなった。

②所要時間の最短群は、回答の一貫性の点からは最も主要回答との一致度が高い群であった。このことは、この群は回答に際して迷いが少ないこと、そして明確な回答基準を持ち、それが調査を通して一貫していたことを示唆する。回答所要時間の短さは不良回答の1つの基準であったが、性格検査についてはそれが当てはまらない。

③中央カテゴリ集中傾向を示す群に関しては、回答の一貫性は方法1では低い、方法2ではその他の群との差は小さく、主要な回答傾向とは逆の回答が少ないという特徴も見られた。この群の回答パターンの特徴は不良回答を予想させるが、判断に迷うケー

スでの回答留保（中央のカテゴリで回答）
という意味での「手抜き回答」の可能性は
あるが、全体としては、信頼性は低い。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者に
は下線）

〔雑誌論文〕（計2件）

- ①加藤岳久・中澤優美子・漁田武雄・山田文
康・山本匠・西垣正勝、本人認証技術にお
けるユーザの性格とセキュリティ意識との
相関に関する考察 情報処理学会論文誌,
Vol. 52, No. 2, 2537-2548, 2011、査読あ
り
- ②早川敬一・山田文康・高嶺一男、アンケー
トにおける不良回答識別手法に関する新
たな試み—SD法質問形式を用いて—、マ
ーケティング・リサーチャー、No. 112、
p41-50、2010、査読あり

〔学会発表〕（計5件）

- ①山田文康・早川敬一・高嶺一男、SD法調査
における項目提示の無作為化による回答
への影響、日本社会情報学会 2012年研究
発表論文集、p231-236、2012.9.16、群馬
大学
- ②早川敬一・山田文康・高嶺一男、「不良回
答」がSDデータの分析結果に及ぼす影響
について、日本社会情報学会 2012年研究
発表論文集、p225-230、2012.9.16、群馬
大学
- ③山田文康・杜勝男・関口茉莉・早川敬一・
高嶺一男、SDデータに対する「不良回答」
識別方法の一般的質問項目への拡張、日本
社会情報学会2011年研究発表論文集、
p427-432、2011.9.11、静岡大学
- ④早川敬一・杜勝男・高嶺一男・山田文康、
SDデータに対する「不良回答」識別のため
の新たな手順の提案、日本社会情報学会

2011年研究発表論文集、p423-426、

2011.9.11、静岡大学

- ⑤山田文康・早川敬一・高嶺一男、アンケー
トにおける「不良回答」の回答特性と分析
結果に与える影響に関する研究、日本社会
情報学会 2010年研究発表論文集、p227-230、
2010.9.5、長崎県立大学

6. 研究組織

(1) 研究代表者

山田 文康 (YAMADA FUMIYASU)
静岡大学・情報学部・教授
研究者番号：40158217

(2) 研究分担者

漁田 武雄 (ISARIDA TAKEO)
静岡大学・情報学部・教授
研究者番号：30116529

(3) 連携研究者

中村 隆 (NAKAMURA TAKASHI)
統計数理研究所・教授
研究者番号：20132699