

匿名性保証のための質問票構造の動的変更

岩井 淳

群馬大学 社会情報学部

1. はじめに

社会調査において調査対象者の負担の軽減は重要な問題であり、これには協力に要する時間上、精神労働上のコストを減らすだけでなく、回答内容の流出などプライバシーの保全に関わる不安を軽減することも含まれている。

本研究では、特に無記名の調査でも結果的に調査対象者の匿名性が守られない場合がある問題に着目し、この問題を回避できる RDB を中心とした調査用システムを設計し、プロトタイプを試作した。利用場面として想定するのは大学組織や企業組織など小規模組織で行う社会調査である。提案システムは、質問票の内容を質問項目ごとに分け、調査対象者が自分を特徴づける可能性の高い項目を他の部分から切り離すことを可能にする点に特に特徴がある。本研究では、ソフトウェアとしての構築だけでなく、人間組織によるその運用プロセスの問題を含めてシステムの実装上の問題を整理する。

以下の構成は次の通りである。準備として、まず匿名性が失われる事例と問題の構造を説明する (2.)。その上で、同問題に対する対処法として本研究で提案する質問票構造の動的変更の枠組を説明する (3.)。

2. 匿名性が失われる事例と問題の構造

最初に準備として、匿名性が結果的に失われる問題を例を用いて整理する (例 1)。また、複数回に分かれた調査資料間の関係の問題を別例を用いて説明し (例 2)、実現に要する技術の観点では両者が同じ構造をもつことを説明する。

例 1: SNS 利用に関する調査に協力した A は、政治問題に特定の立場から発信を続けたことを記述した後で、主な使用 OS と Web ブラウザを問われた。A は一般的でない OS を使用しており、

そのことがある程度知られていた。正しい記述は匿名性確保の上で問題があり、一人その部分を空欄として残すことにも同様のリスクがあった。回答全体の破棄も考えたが、既に回答した部分の内容的価値を考えた A は、同質問項目で偽って一般的な OS 名を記述した。

例 2: ある組織で、1 年間に計 n 回の無記名調査を行った。結果として n セットの資料を高回収率で得たが、回答票間対応の情報は失われたため、その点で分析上の制約が残った。

例 1 は、正しく回答すると匿名性が結果的に失われる例であり、問題の質問項目が他の部分と同一質問票上にあることに一因が求められる。全回収票で同質問項目が切り離されれば、A の匿名性が維持される事情があるためである。例 2 は複数回に分かれた調査資料間の関係の問題であり、問題回避に求められる技術という点では、例 1 と同様、質問項目間の連続性を操作する技術に関わる。例 1 は 1 質問票内で特定項目を独立化させられなかった問題、例 2 は複数の質問票間で項目を関連付けられなかった問題と位置づけられる。

3. 質問票構造の動的変更の枠組

例 1, 例 2 の問題への対応として、複数表間の関係を SQL で容易に指定できる RDB の特徴を活用し、質問項目間の連続性をこの表間の関係として表現する枠組みを提案する。特定質問項目を他の部分から容易に独立化でき、同時に複数の質問票間で項目を関連付けられるようになる。ここで連続性については、システムは以下の仕様を満たすとする: A) 全質問項目を項目ごとに表にまとめる。B) 原則として、異なる質問項目間で、同一回答者による回答レコードの対応付けが可能。C) ただし、調査対象者は任意の質問項目を、対応づけ不可能な独立項目に設定できる。以下、SQL で RDB とアクセスする PHP と連携可能な web サーバの利用を念頭におき、各調査協力者は個別にこのシステムにアクセスして調査項目に回答するものとする。

調査技法の手続きは以下の通り: 1) 調査対象

Dynamic Change of Structure of Query Sheet for Guaranteeing Anonymity

Atsushi IWAI

Gunma University, Faculty of Social and Information Studies

者への説明（各対象者は無作為にカードを選択、固有のアクセス ID と password を得る）、2) 調査用システムの公開、3) 調査対象者の Web アクセス（RDB と連携する PHP コードが各質問を表示。回答は項目ごとに独立した表として RDB に記録される。各表は、主キーを別として、アクセス ID と回答内容を記録する 2 列からなる）、4) 回答内容に対する分析。匿名性の維持と回答の連続性が両立するのは無作為に選択したアクセス ID を連続使用する規則による。また、調査対象者が特定の質問項目の独立項目化の操作をするのは上記 3) の段である。

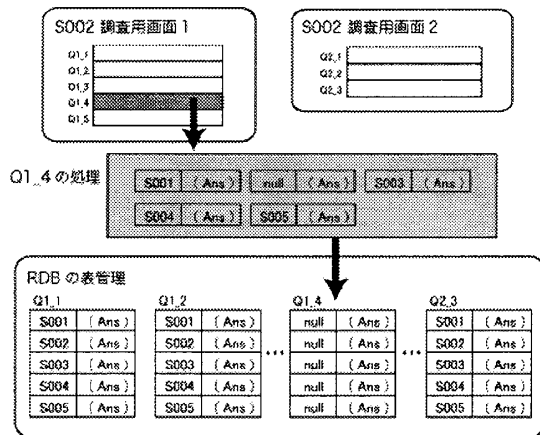


図 1 調査対象者による情報隠蔽要請の処理

図 1 は、同システムで調査対象者により独立項目化の要請がなされ、自動処理される手続きの説明である。簡単のため、調査対象者は S001 から S005 の 5 名、調査は 2 度で、1 度目は Q1_1 から Q1_5 の 5 設問、2 度目は Q2_1 から Q2_3 の 3 設問としている（“Ans” は各回答内容の略記）。この例では、S002 が Q1_4 の独立項目化を要請（box に check）している。このとき、S002 の Q1_4 用データの第 1 列は null となる。また、同時に他の 4 調査対象者の Q1_4 用データ第 1 列は null に置換され Q1_4 の表と他表との関係づける情報は失われる（Q1_4 が独立項目となる）。S002 の Q1_4 以外に対する回答内容が知られることはなくなるため、S002 は Q1_4 に事実に基づいた回答ができるようになる（例 1 の問題への対処。Q1_4 以外の全項目間で、レコード間の対応確認が可能であるため、例 2 の問題も合わせて解決される。）

なお、例 2 の問題の観点から対処法の比較として、全質問票（全質問項目）を個人別に封書にまとめ（封書内の項目には封書ごとの個別の番号を一律に付す）無作為配布する方法を考え

る。この場合も本提案技法と同様の結果が得られるが、本提案技法は、1) 調査開始後に新たな調査を加えられる点、2) 例 1 の問題を合わせて解決する点で（独立項目扱いの要請を受けてから、他の調査対象者の回答をその内容を確認せずに処理する手続は手作業では実現しにくい）、より有用である。

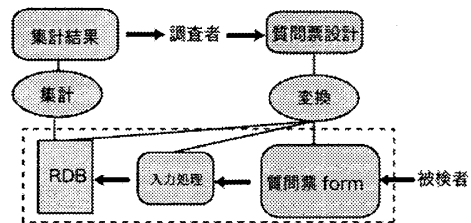


図 2 調査用システムとその利用

本研究では、上記調査システムのプロトタイプを作成し、機能的には適切に動作することを確認した。図 2 は同システム（破線部）利用の概念図であり、調査者が質問票を設計して集計結果を得るまでの過程はシンプルである。ただし、通常の RDB 処理で用いない情報隠蔽の手続きを含んでいる。例えば、RDB では全レコードを区別する必要があるため実際には Q1_4 のように全レコード内容を同一とはできず別変数を用いる必要があるが、この別変数が被験者特定の情報となってはならない。このため別変数とレコード内の他の情報との対応づけを適宜無作為に改める手続きを含めるように設計した。

4. 結語

本研究では、無記名の社会調査でも結果的に調査対象者の匿名性が守られない場合がある問題を検討し、同問題を回避する調査技法を提案してプロトタイプとなるシステムを作成した。提案技法は RDB を中心とするアプローチに特徴があり、シンプルで実現に要するコストが小さい。RDB を用いた社会分析方法には歴史資料に関する[1]の提案等が既にあるが、本研究は被験者の匿名性に着目した点に特徴がある。

謝辞

本研究の一部は、日本学術振興会科学研究費補助金（課題番号：18700246）の助成による。

文献

[1]中里英樹. 「リレーショナル・データベースによる定型データの作成」, 理論と方法, pp197-212, 2004.