

令和 4 年 6 月 7 日現在

機関番号：16201

研究種目：挑戦的研究（萌芽）

研究期間：2018～2021

課題番号：18K18604

研究課題名（和文）アンケートの問い方を変える：ネットワーク科学を活用した自由記述式の統計分類法

研究課題名（英文）graph-based surveys

研究代表者

青木 高明（Takaaki, Aoki）

香川大学・教育学部・准教授

研究者番号：30553284

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：意思決定に際して「人々の意見や考え、思いを聞きたい」という社会ニーズがある。特に自由記述式アンケートは、多様な意見を集めることができる手法であるが、自由記述回答の分類を分析者が人手で処理する必要があり、ボトルネックとなっていた。本課題では回答者自身に、自分の意見と他人の意見との「似ている」「似ていない」という意見間の関係性を判断してもらうことで、機械学習による統計的分類を実現し、香川大学教育学部卒業生アンケート等において提案法の有効性を示した。

研究成果の学術的意義や社会的意義

近年、インターネットを介したオンライン調査によって、1万人以上の人々から多様な意見を収集することが可能になった。この手法を活用することで、「どうして研究者を目指したのですか？」「大学で学んで良かったことはなんですか？」といった「なぜ？」「どうして？」を問いかける大規模な自由記述式アンケートを実施し、その大量の回答から意見グループを自動抽出し、統計分析することが可能となる。世論調査・顧客調査などに活用し、より良い製品開発や社会実現のために役立てることが出来る。

研究成果の概要（英文）：Why did you become a scientist? You may have your own story, and such individual answer can be stimulating and informative. But, how do we summarise the answers from many people? This is a major scientific issue in open-ended surveys called coding. To perform statistical analysis, individual responses must be classified into several groups, capturing the underlying semantics that respondents expressed. This manual task can be difficult, subjective, and time-consuming, and this bottleneck has limited the utility of large-scale surveys. We proposed a network-based survey method in which responses are automatically classified in a statistically principled manner, and demonstrated it using a survey taken by graduates of a university. The method allows us to ask open-ended questions to large populations, such as ``What is the strongest point of our service?'' or ``Why did you buy this product?'' This would bring substantial benefits for society and industry.

研究分野：ネットワーク科学

キーワード：アンケート分析 コミュニティ分解 ネットワーク科学 自由記述式アンケート 機械学習

1. 研究開始当初の背景

○ 研究背景：社会基盤としてのアンケート調査

意思決定に際して「人々の意見や考え、思いを聞きたい」という社会ニーズがある。アンケート調査は社会調査の基盤技術として、社会学や心理学等の学術調査のみならず、世論調査・消費者アンケートや市場調査等、政府・大企業から個人・商店に至るまで広く社会的に利用され、新商品の開発や政策決定に役立てられている。

アンケートによる質問には選択肢回答方式と自由記述回答方式の2つがある

選択肢式については、

- ✓ メリット：「賛成/どちらともいえない/反対」のように回答を選択肢として標準化することで、結果に対する統計分析が適用しやすい。
- ✓ デメリット：選択肢による限定は、対象者の考え・意見を制限する可能性がある。

自由記述式では

- ✓ メリット：バラエティに富んだ生の声を集めることができ、驚きや発見に満ちた意見から課題に対する新しい知見が導かれる事もある。
- ✓ デメリット：自由記述回答を統計分析し結論を導くためには、分析者が文章を精読し、多様性を失わないようにしつつ、少数のグループに分類する(コード化)などの処理が必要となる。一般にアンケート調査においては、これらのメリット・デメリットの他、対象の特性やこれまで得られた知見等を考慮し、2方式を組み合わせた適切な調査票の設計が不可欠である。

○課題：大規模アンケートでは自由記述式が利用困難

近年、インターネットを介したオンライン調査によって1万人を超える回答を集めることも技術的に可能になった。しかし大規模調査において、自由記述回答を分析するためには、膨大な量の回答を分析者が人手で処理する必要があるため、実質的に利用が不可能となっている(コード化のボトルネック問題)。

そのため選択肢式が必ずしも適さないような質問、例えば「なぜ研究者を志したのか?」「次の選挙で一番重視している点は何か?」といった質問に対して、不特定多数の意見を事前に予想できない場合でも、ありそうな選択肢を設定し、回答者に選択させる必要があった。

従来研究においては、少人数を対象に自由記述式で事前調査を行う対策も取られてきた。しかし追加コストや少数サンプルの問題が残る。また近年著しく研究が進められている、テキストマイニングを使った機械分析もあるが、自然言語処理にはいまだ信頼性に課題が残る。

2. 研究の目的

「自分の言葉で書かれた意見」を統計的に分析する手法の開発

自由記述式の大規模調査という『絵に書いた餅』を現実にするのが本研究の目的である。つまり、「賛成/どちらともいえない/反対」といった標準化された回答の統計分析を越え、回答者の「生の意見」という非標準データの統計分析法を開発し、実践する。

3. 研究の方法

○意見ネットワークの構築による回答者判断に基づく統計的分類手法

提案法では、(1)通常自由回答方式の質問と回答を実施し、(2)過去に得られた回答リスト

を10個程度提示し、自分の意見と「似た/似ていない」を選択肢回答させる(前ページの図1)。この結果、各人の意見が「似た・似ていない」の関係で繋がったネットワークを構築する。ネットワーク科学のコミュニティ検出法を応用し、似た意見のグループを自動抽出し、多様な自由記述を複数グループに分類する。この分類は名義変数として選択肢回答と同様に、クロス表分析や回帰分析などの統計解析が可能である。

似た意見のグループ抽出については、ベイズ統計に基づく統計推論として定式化する。グループ数の決定についても、予測誤差等を基準としたモデル選択手法(川本&樺島, Scientific Reports, 2017)により決定することで統計的に根拠ある分類を推定できる。

○提案法の利点

回答文からその意図を正しく読み解き、さらに少数のグループに分類することは、熟練した研究者にとっても容易ではない。開発のポイントは、回答者自身に内容を評価してもらい、その結果をネットワーク(グラフ)の形としてまとめたことにある。これにより、

1. グラフ分割という数学的視点から、統計推論による定式化とアルゴリズムベースの分類が可能になった。
2. 回答者からの視点による分類が可能になった。分析者個人による判断ではなく、また自然言語処理技術が依拠する辞書・例文データによる不特定多数の視点でもない。回答者達の投票に基づく「民主的」な分類を可能にしている。

4. 研究成果

香川大学教育学部卒業生アンケート調査に対して、提案手法を適用した。「社会に出て改めて、大学で学んでよかったことは何ですか?」という質問に対しては、図1のように「教養・友人」の意見(緑色の丸)、「資格・専門知識」の意見(紫色の丸)、「自己探求・教室外での学び」の意見(黄色色の丸)の3種類の意見を抽出することができた。緑・紫・黄色の3色が明確な丸はグループの典型的意見を示し、一方で中間色はグループ間にまたがる中間的意見を示している。多様な意見を3グループにまとめ上げると同時に、その多様性を可視化することができた。これらの成果を研究論文として出版した(Democratic classification of free-format survey responses with a network-based framework, Tatsuro Kawamoto and Takaaki Aoki, Nature Machine Intelligence, vol.1, 322-327 (2019))。

Q. 社会に出て改めて、大学で学んでよかったことは何ですか?

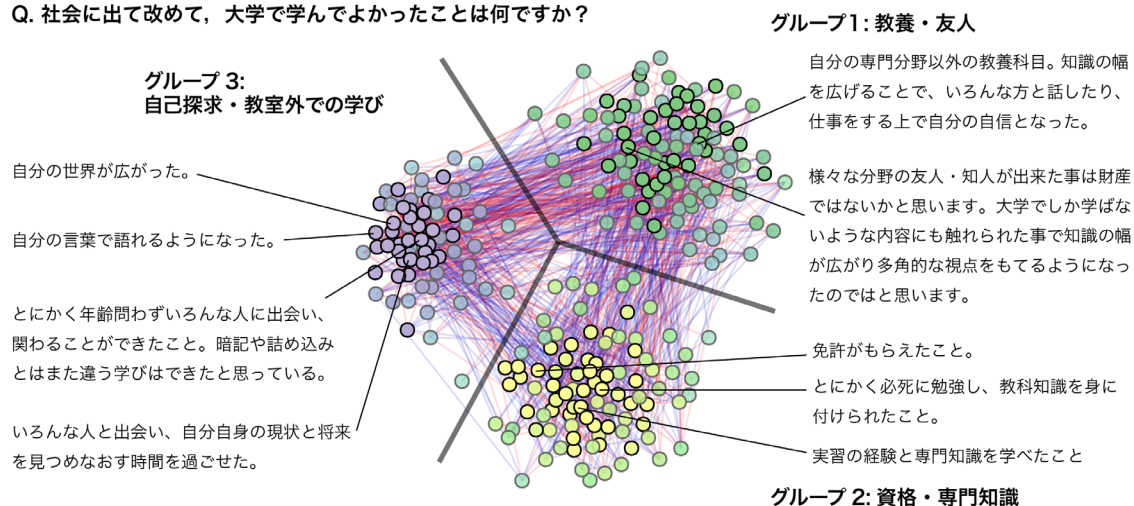


図1: 2016年香川大学教育学部卒業生アンケート調査での分析例

また現在のコロナウイルス感染拡大の状況のもと、人々が何を悩み、不安に思っているのかと

いう点について、心情調査アンケートを分析した。社会状況の外面的状態は後になって把握すること可能であるが、人々の思いや心情はその時々、収集・記録しなければ失われてしまう情報である。よって定期的に、1000人以上の全国の回答者(代表的サンプル)を対象に心情調査アンケートを実施した。

人々の意見は大きくは6-9個程度に分類することができた。調査時期・感染拡大状況に応じて典型的意見が変化していた。選択肢式アンケートでは事前に回答例を予想し、選択肢として準備しておく必要があるが、今回の心情調査アンケートの場合には、そのような事前予想は困難であったと思われる。そのため自由記述式アンケートで実施した意義が認められる。これらの成果を研究論文として出版した (Graph-based open-ended survey on concerns related to COVID-19, Tatsuro Kawamoto, Takaaki Aoki & Michiko Ueda, PLOS ONE, vol. 16(8), 0256212 (2021).)

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Kawamoto Tatsuro, Aoki Takaaki	4. 巻 1
2. 論文標題 Democratic classification of free-format survey responses with a network-based framework	5. 発行年 2019年
3. 雑誌名 Nature Machine Intelligence	6. 最初と最後の頁 322 ~ 327
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s42256-019-0071-y	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kawamoto Tatsuro, Aoki Takaaki, Ueda Michiko	4. 巻 16
2. 論文標題 Graph-based open-ended survey on concerns related to COVID-19	5. 発行年 2021年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0256212
掲載論文のDOI (デジタルオブジェクト識別子) 10.1371/journal.pone.0256212	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 川本達郎, 青木高明, 上田路子
2. 発表標題 グラフベースの自由記述式アンケートシステムを用いた新型コロナウイルスについての心情調査研究
3. 学会等名 日本物理学会 第76回年次大会
4. 発表年 2021年

1. 発表者名 川本 達郎
2. 発表標題 グラスマン代数を用いたキルヒホッフ恒等式の導出
3. 学会等名 日本物理学会2019年秋季大会
4. 発表年 2019年

1. 発表者名 Chihiro Noguchi, Tatsuro Kawamoto
2. 発表標題 Evaluating network partitions through visualization
3. 学会等名 NetSci2019
4. 発表年 2019年

1. 発表者名 Chihiro Noguchi, Tatsuro Kawamoto
2. 発表標題 Robustness of spectral clustering for networks with an overlapping community structure
3. 学会等名 NetSci-X Tokyo
4. 発表年 2020年

1. 発表者名 Ryutaro Hashimoto, Tatsuro Kawamoto
2. 発表標題 Numerical performance assessment for inference of block models with multiple edge labels
3. 学会等名 NetSci-X Tokyo
4. 発表年 2020年

1. 発表者名 青木高明
2. 発表標題 グラフベースの自由記述式アンケートの統計分類法
3. 学会等名 第14回京都算学会
4. 発表年 2018年

1. 発表者名 川本 達郎
2. 発表標題 voteclustering: A network approach to free-response surveys
3. 学会等名 Network Science Workshop
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

投票クラスタリング https://ja.voteclustering.org/
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	時岡 晴美 (Tokioka Harumi) (10155509)	香川大学・教育学部・特命教授 (16201)	
研究分担者	川本 達郎 (Kawamoto Tatsuro) (10791444)	国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究員 (82626)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------