

令和元年 7 月 10 日

全く新しい大規模な自由記述式アンケートの方法を開発

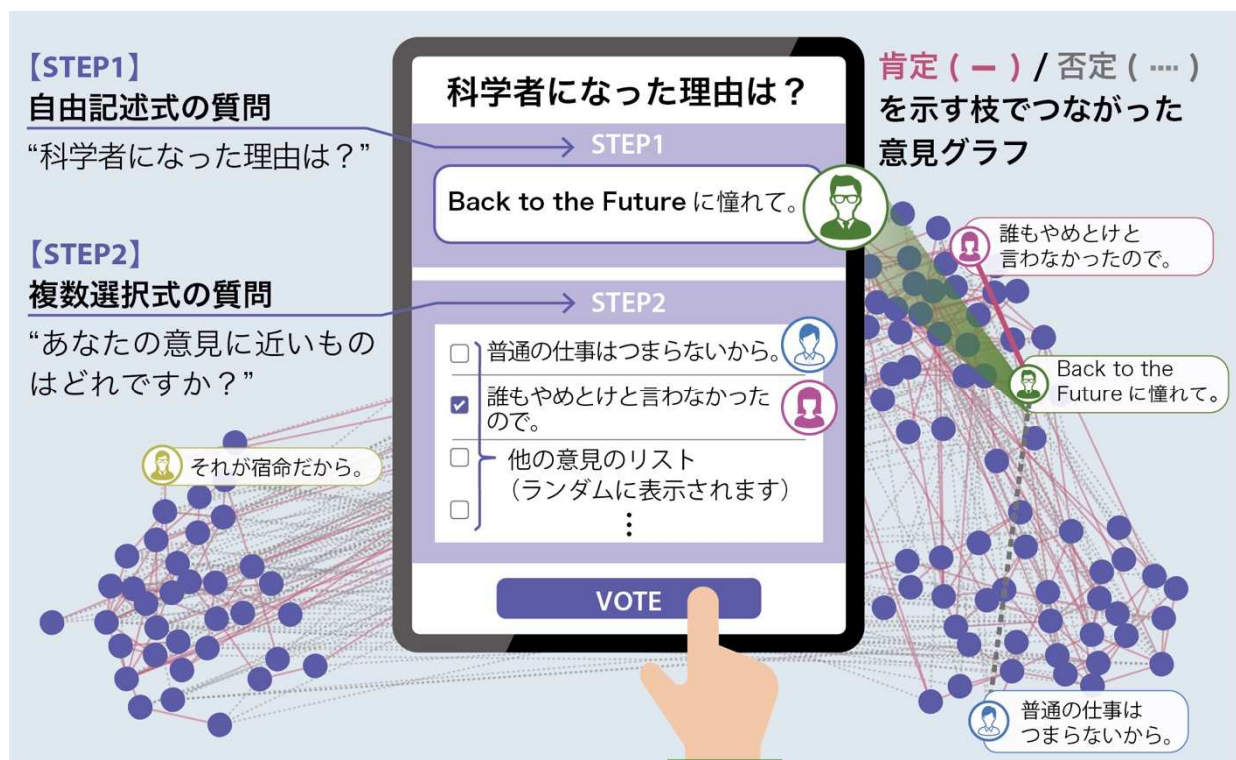
－ 自由記述回答文からの情報集約を大幅に効率化 －

研究のポイント

- ・ 自由記述式アンケートの回答文を、自然言語処理を使わずに効率的に集約する手法を開発
- ・ アンケートの意見集約という問題を、グラフ分割という数学の問題に変換
- ・ これまで困難だった大規模な自由記述式アンケートを実現

概要

川本達郎 研究員(国立研究開発法人 産業技術総合研究所 人工知能研究センター)と、青木高明 准教授(国立大学法人 香川大学 教育学部)の研究チームは、全く新しい自由記述式アンケートの方法「投票クラスタリング」(英語名: voteclustering)を開発した。意思決定において「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。しかし従来のアンケート分析では、自由記述回答の分類を分析者がすべて読んで人手で処理する必要があり、大規模な実施が難しかった。研究チームは「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(投票クラスタリング)」というアンケート手法を開発し、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。この手法によって、「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?」といった質問を1万人以上の人々に投げかけ、その大量の回答から意見グループを効率的に抽出し、統計分析することが可能となる(図1)。なお、この技術の詳細は、2019年7月9日(イギリス時間)に Springer Nature から出版される機械学習分野の雑誌 Nature Machine Intelligence に掲載された (DOI: [10.1038/s42256-019-0071-y](https://doi.org/10.1038/s42256-019-0071-y), SharedIt: <https://rdcu.be/bJpUy>)。 _____ は【用語の説明】参照



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)

研究の背景

広く人々の声を聞き、そのニーズを捉えることは商品開発・政策決定などに需要がある。そのため、顧客アンケート・世論調査などは、我々の社会のいたるところで実施されている。特に自由記述式アンケートは、選択肢に限定されることなく、人々が持っている多様な意見を、彼ら自身の言葉で集めることができる有効な手法である。しかし大量の意見を要約・分析するためには、分析者が回答文を手手で分類する必要がある(回答文のコード化処理)、大規模調査への活用が実質困難となっていた。代替として「よかった」「よくなかった」「どちらともいえない」などのレーティング評価や、複数の選択肢からの回答を選択する調査が多いが、得られる情報が限定され、人々の生の声を集めることができないという問題がある。自由記述回答を自然言語解析技術で処理する手法も提案されているが、必ずしも正しく回答の意図を汲み取ることが保証されないという問題がある。そのため、自由記述回答の内容を読み解き、素早く分類する手法の開発が求められていた。

研究の経緯

産総研の川本 達郎研究員は、社会における多様な意見を抽出し、理解しやすい形で集約することが、高いレベルの民主主義社会を実現するために重要であるとの考えから、日常生活における価値観や意見を観測するための重要な手段の一つである大規模な自由記述形式アンケート手法の研究を進めてきた。2016年からは、香川大学教育学部の青木高明准教授との共同研究を開始した。研究チームは、香川大学教育学部卒業生に対するアンケート調査等を実施し、手法の改善と実証を進めてきた。今回の論文発表はこれらの研究成果をまとめたものである。

本研究は、日本学術振興会科学研究費補助金(特別研究員奨励費「ネットワーク上のコミュニティ検出問題への物理学的アプローチ」14J11203・挑戦萌芽研究「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」18K18604)の支援のもとで行われた。香川大学教育学部卒業生に対するアンケート調査は、同学部の時岡 晴美教授の協力の下、香川大学教育学部および同学部同窓会(松楠会)の協力・支援を受けて実施された。

研究の内容

今回開発したアンケート手法では、回答者に対して、ランダムに抽出された他の回答者の回答が一定数提示される。各回答者はそれらが自分の意見に合致するかを判定し、もしも自分の意見が提示された回答と完全に合致しない場合には、自身の回答を記入する。このようにすることで、個々の回答意見を、回答者たちの判定(一致するか否か)でつないだグラフ(ネットワーク)データが得られる。このグラフを機械学習アルゴリズムの一つであるグラフ分割アルゴリズムによって処理することで、意味として類似した回答をまとめ上げることができる。

具体例の一つとして、開発したアンケート手法の有効性を検証するために2016年に実施した、香川大学教育学部卒業生に対するアンケート調査の回答を解析した結果を図1に示す。例えば、「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、「教養・友人」に関する意見(緑色の丸)、「資格・専門知識」に関する意見(黄色の丸)、「自己探求・教室外での学び」に関する意見(紫色の丸)の3種類の意見のグループを抽出することができた。今回開発した手法によって得られた分類結果は、分析者の主観的分類ではなく、回答者たちの主観を反映した集約になっている。

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？

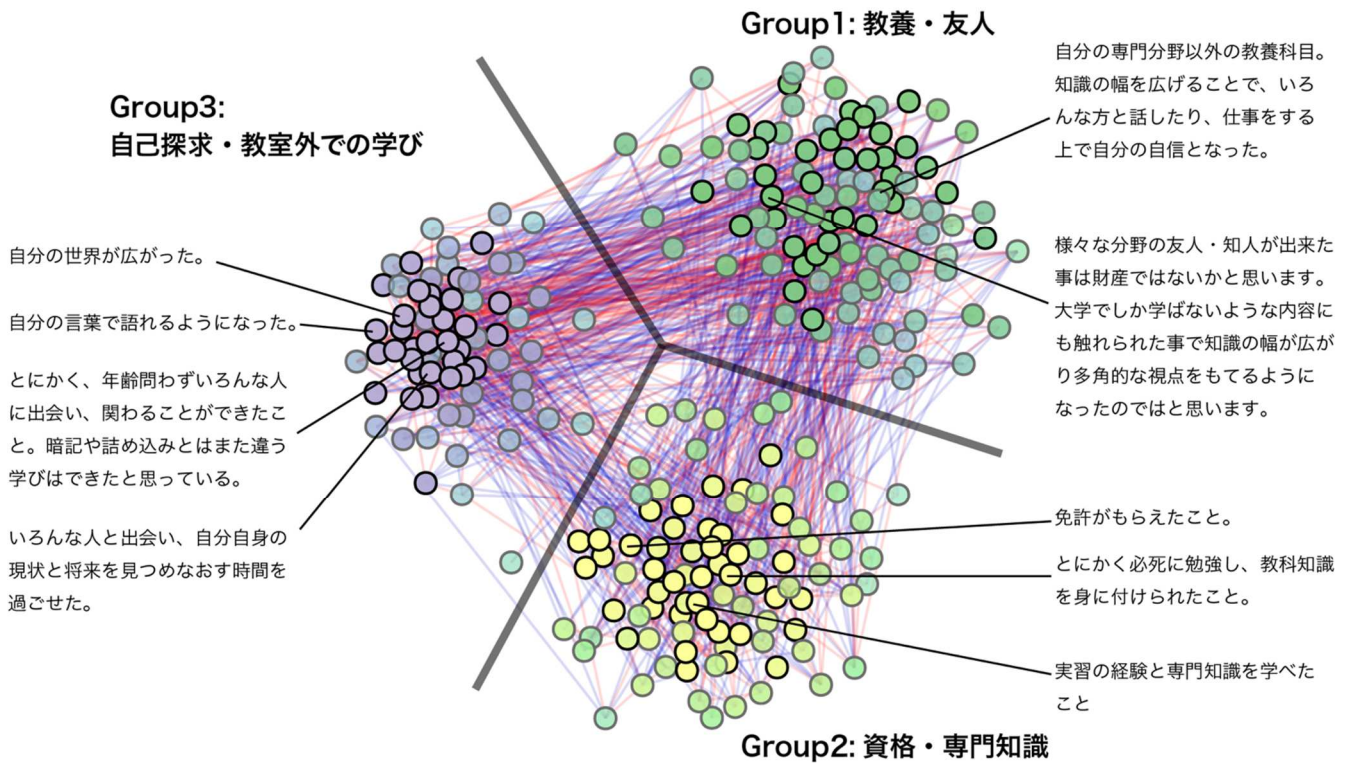


図1 2016年香川大学教育学部卒業生アンケート調査での回答文の分析例

従来の自由記述式アンケートを困難にしていたのは、分析者がすべての回答文を読んで処理する必要があり、大規模に実施することが難しいという点であった。今回開発した方法は、その作業を劇的に効率化する。意見を自動的にグループ化した後は、アンケートの結果が解釈可能なものである限り、少数の代表的な回答を読むだけで各グループの意見を読み取ることができる。

この方法は、アンケート中に他人の意見を参照させることが大きな特徴である。このことは、情報を与えない状況下で回答を集めることことを重要視するような実験・調査には適さないが、回答者により深く考えさせ、多様な意見を抽出することにつながるというメリットもある。社会における多様な意見を抽出し、理解しやすい形で集約することは非常に重要であり、今後、様々な場面での活用が期待される。

今後の予定

今後は、広く大規模自由記述式アンケートの機会を提供し、実証実験を重ねていく。これを通して、より高精度なデータ処理方法の研究開発を進め、実サービス化を可能にすることを目指す。

論文情報

掲載誌: Nature Machine Intelligence 2019 年 7 月 9 日(イギリス時間)公開

論文タイトル: Democratic classification of free-format survey responses with a network-based framework

著者: Tatsuro Kawamoto and Takaaki Aoki

問い合わせ先

[研究内容について]

国立大学法人 香川大学 教育学部

准教授 青木 高明

〒760-8521 香川県高松市幸町 1-1 香川大学教育学部

TEL: 087-832-1595

E-mail: aoki@ed.kagawa-u.ac.jp

[プレスリリースについて]

国立大学法人 香川大学 教育学部総務係

〒760-8521 香川県高松市幸町 1-1

TEL:087-832-1405 FAX:087-832-1418

E-mail: lsoumut@jim.ao.kagawa-u.ac.jp

【用語の説明】

自由記述式アンケート

設問に対して、選択肢の中から該当するものを選ぶのではなく、自由な言葉で意見を回答させるようなアンケート。

グラフ分割

グラフ(ネットワーク)データのなかから特徴的な構造を抽出する、機械学習の問題。

グラフ分割アルゴリズム

グラフ分割を実行するための手続き。