Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

Web アンケートにおける不真面目回答の ChatGPT を用いた自動分類

畑中健壱[1]　山﨑郁未[1]　中村聡史[1]

概要：Web アンケートは手軽にデータ収集ができ便利であるものの，設問に対して真面目に回答しない回答が多く集まる．特に自由記述においては「わからない」や「特になし」など不真面目回答をする人が集まる問題がある．このような問題に対して，設問の順序や回答欄のデザインなどに工夫がなされている．その工夫の評価指標として不真面目回答率が用いられているが，評価者が全ての自由記述回答を手作業で判断するため，膨大なデータを扱う Web アンケートでは非常に手間であり時間がかかる．ここでアンケートの質問とそれに対する回答は一種の対話と捉えることができる．そこで文の意図や内容を理解できる対話型 AI，ChatGPT を用いて回答の自動分類をできるのではないかと考えた．本研究では，ChatGPT を用いた不真面目回答分類の基礎検討として，これまで行ってきたアンケートデータを用いて分類精度や効率的なプロンプトについて検討を行った．実験の結果，回答に対して点数を付与する手法が最も精度高い結果となった．一方，F 値は十分ではなく，判定における課題が残った．

キーワード：Web アンケート，回答分類，不真面目回答，ChatGPT

1. はじめに

Web アンケートは紙ベースのアンケートに比べ，手軽に多くの回答を集めることが可能である[1]．その中で自由記述形式の設問は，回答者の多様な回答を得られることができる[2][3]ためアンケートにおいて必要不可欠である．そのため，自由記述形式の設問は，教育，医療，市場調査や社会科学研究など，多岐にわたる分野のアンケートに取り入れられている．しかし，自由記述形式の設問は，選択式の設問などと比べ回答に時間がかかる[4][5]ことから，回答するための負担が大きい．そのため，設問に全く答えない，あるいは考える必要のない回答をすることで負担を減らそうとしている回答者がおり[6]，選択形式の設問よりも不真面目回答が多く集まってしまう[2]．例えば，Holland ら[7]は，自由記述形式の設問で，そもそも何も回答しない無回答や，「わからない」，「fdjkgfg」，「xxx」，「ただ何となく」などの意味を持たない回答が集まったと述べている．

このような不真面目回答や無回答の問題に対して，アンケートのデザイン，特に自由記述の設問文や設問の位置を工夫することにより，不真面目回答を減らす研究が行われている．Zuell ら[8]は自由記述の設問文で動機づけ文章を追加することにより，動機づけ文章なしの場合と比べ，無回答を減少させることを明らかにしている．また Yamazaki ら[9]は，自由記述設問を最後に配置する方が，最初に配置するよりも不真面目回答の割合が少なくなることを明らかにしている．

こうしたアンケートデザインの有効性の評価に，不真面目回答率が用いられる．不真面目回答率は，不真面目回答がその自由記述設問にどのくらいの割合で含まれているかを表すものであり，複数名の評価者が 1 つ 1 つの回答に対して，真面目か不真面目かを判断し算出する．そのため，大規模調査を行う Web アンケートは不真面目回答を分類するために，時間と労力がかかる．また真面目か不真面目かの判断には主観性を伴うため，分類の一貫性を保つのが難しく，アンケートテーマによっては評価者の事前知識が必要となる場合がある．なお，アンケートデザインに限らず，一般のアンケートにおいてもこうした不真面目回答を判定して除外することは重要である．そこで，これらの問題を解決する方法として，機械学習などによる不真面目回答の自動分類が挙げられる．

不真面目回答の基準は研究によって異なっており，Don’t Know 回答（DK）を問題視する研究[10]や，無回答（Item nonresponse）を問題視する研究[8]，DK 回答および意味をなさない回答（non-substantive response）を問題視する研究[6]，質問に対する答えが伴っていない回答と意味をなさない回答（Non-earnest response）を問題視する研究[9]など，様々である．このように基準が研究によって異なると単純に比較することができない.そのため，不真面目回答率を他の研究と比較することのできる共通の基準を作ることが重要である．

ここでアンケートの不真面目回答は，質問に対して回答がずれているとも捉えることができ，これは日常的な会話の中における質問への回答がずれている問題と似ている．例えば，話を聞いていない場合や真面目に回答する気がない場合の回答は，その会話の質問に対する回答は不適切な回答であるといえる．ここで ChatGPT[11]などの対話型 AI は，大量のテキストデータを使ってトレーニングされた自然言語処理モデルを用いているものであり，質問への自然な回答を得意としている．つまり，ChatGPT は不真面目回答のような質問に対する不自然な回答を検知することができるのではないかと考えられる．また ChatGPT は特定の

1 明治大学
Meiji University

Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

訓練データに依存することなく実施することができるため，異なる研究や調査に対しても一貫した基準で不真面目回答率を定量化することが期待できる．

そこで本研究では，まず自由記述設問の不真面目回答を，ChatGPT を用いて自動分類することが可能であるかを調査することを目的とし，様々なプロンプトで ChatGPT を用いた不真面目回答自動分類ができるのか，また効果的なプロンプトおよび自動分類方法についても検証を行う．

2. 関連研究

2.1 自由記述設問の回答行動

Web アンケートにおける不真面目回答に関する研究は数多く行われている．Reja ら[2]は，オンラインアンケートにて選択形式の設問と自由記述形式の設問で回答の比較を行ったところ，選択形式の設問より自由記述形式の設問の方が，欠損データが多く存在することを明らかにしている．Holland ら[7]は，回答するアンケートのトピックの関心度が回答にどう影響するのか調査を行った．その結果，トピックへの関心度が高い人は回答の質が高く，関心がない人や低い人は自由記述形式の設問で無回答が多くなることを明らかにしている．また Ronggang ら[12]は，単体の自由記述と選択設問に対する理由を答えてもらう自由記述の 2 種類を含みアンケートにより実験を行ったところ，75%以上の人がどちらの自由記述にも回答をしないことを示した．Schmidt ら[13]は自由記述の設問が後ろにあるほど，解釈可能な回答をする度合いが有意に低くなることを明らかにしている．Galesic ら[14]は，質問が後ろになると質問開始直後より回答時間が短く，自由記述の回答文が短いことを明らかにしている．このように自由記述設問では，真面目回答ではない回答が多く集まる．またそれを評価するのは手作業で行われる．そのため本研究ではこうした不真面目回答を自動で分類することを目指す．

2.2 自由記述設問の工夫

自由記述設問において，より良い回答を得るためにアンケートデザインに関する研究が様々行われている．Smith[15]は，自由記述設問の回答スペースを広く取ることで，回答が長くなり，実際の口頭表現に近い回答が得られるようになったことを明らかにしている．Emde ら[16]は，大学生を対象としたアンケート調査において，無回答を減らしつつ長い回答を得るため，回答に応じてテキストボックスのサイズの大きさを自動で拡張する回答欄を提案した．また，最初の自由記述設問で回答した文字数に応じて，2 つ目の自由記述設問でカスタマイズした回答欄を割り当てるアンケートを提案した．実験の結果，回答に応じてカスタマイズした回答欄を割り当てることで，無回答を減らし回答の質が向上することを明らかにしている．山﨑ら[17]は，自由記述設問の順番および，テキストボックスの関係について調査を行った．その結果，自由記述設問を最初に回答，かつテキストボックスが大きいと，離脱率が最も高くなる傾向があることを明らかにしている．このように自由記述設問の工夫は様々行われており，その評価指標として，回答の長さ，無回答，離脱率などが用いられている．本研究では，回答の長さや離脱率など，研究によって左右されない不真面目回答の指標として，ChatGPT を用いた回答分類が利用できないかを検討する．

2.3 自由記述設問のカテゴリ自動分類に関する研究

自由記述設問のカテゴリ分類を自動で行う研究は多数行われている．Kawamoto ら[18]は，回答からネットワークを構築し，そのネットワーク内のコミュニティを利用して回答カテゴリに分類する手法を提案している．従来手法よりも提案手法の方が信頼性の高い分類が可能であることが示唆されている．Schonlau ら[19]は，回答カテゴリの自動分類には限界があるとし，確信度の高いものを自動分類し，低いものを手動で分類する半自動分類を提案している．全自動分類と比較の結果，半自動分類の方が精度が高く半自動分類の有効性を示している．また Gweon ら[20]は，BERT を用いたカテゴリ分類の自動化が可能であるのか調査を行った．サポートベクターマシーンやランダムフォレストなどの一般的な機械学習ベースの分類精度と比較した結果，BERT を用いた分類，特にファインチューニングをした BERT の分類精度が最も高いことが明らかにした．このように回答のカテゴリ分類の自動分類は有効であり，特に BERT のような大規模言語モデルは分類タスクに有効である．本研究でも，大規模言語モデルである ChatGPT[11]を用いて，不真面目回答分類ができるかを調査する．

3. 対象とするデータセット

本研究では，自由記述設問の不真面目回答を，ChatGPT[11]を用いて自動分類できるかを調査するため，2 種類のアンケートデータを用いて検証を行う．本章では 2 種類のアンケートデータについて，アンケートの概要，自由記述設問の内容，不真面目回答分類について述べる．

3.1 運転免許を所持している人向けのアンケート

運転免許を所持している人向けのアンケートは Yamazaki ら[9]が行ったアンケートデータである．このアンケートは Yahoo!クラウドソーシング[21]上で，運転免許を所持している人を対象に行われており，1,000 人（男性 500 人，女性 500 人）に対して行われた．なお有効回答は 979 件であった．

アンケートの自由記述設問の内容と設問ごとの不真面目回答率を表 1 に示す．アンケートは全 17 問で構成され，内自由記述設問は 4 問であった．設問内容はアンケート対象である運転免許を所持している人が，必ず回答できるように構成されており，「特になし」といった回答ができな

表 1 運転免許を所持している人向けのアンケートの設問内容と不真面目回答率

設問番号	設問内容	不真面目回答率(%)
Q-1	主に何のために運転しているか回答してください．普段運転しない方は，なぜ運転免許を取得しようと思ったのか回答してください	2.9
Q-2	主に運転する道の特徴を回答してください．普段運転しない方は，住んでいる家の周辺にどのような道があるか回答してください	6.1
Q-3	運転に苦手意識のある方は，どんなことが苦手か，またはどうして苦手と感じているのか回答してください．運転に自信がある方はどうして自信があるのか回答してください	7.8
Q-4	運転をするときに気をつけていることを回答してください．普段運転をしない方は，運転免許を取得する際に気をつけていたことを回答してください．些細なことでも構いません	5.6

表 2 動物園・水族館に関するアンケートの設問内容と不真面目回答率

設問番号	設問内容	不真面目回答率(%)
Q-1	好きな動物園・水族館を複数回答してください．好きな動物園・水族館がない場合はこれまでに訪問した回数が多い動物園・水族館を回答してください	2.3
Q-2	好きな動物・生き物を教えてください．好きな動物・生き物がない場合は動物園や水族館でよく見かける動物や生き物を回答してください	2.0
Q-3	動物園・水族館で楽しかったことや面白かった経験を書いてください．特にない場合は動物園や水族館でどんな経験をしてみたいかを書いてください	4.6
Q-4	動物園・水族館をリピート訪問したことがあるひとは，その動物園・水族館と理由を回答してください．リピート訪問したことがない人は，リピートしなかった理由を回答してください	12.8

いようになっていた．

不真面目回答率については，「設問に対して答えが伴っておらず，回答そのもので意味を捉えられないもの」という基準で評価者 2 名が不真面目回答分類を行っている．なお，意見が割れていたものは著者らが最終判断をして分類を行っていた．

3.2 動物園・水族館に関するアンケート

動物園・水族館に関するアンケートは我々[22]が行ったアンケートデータである．このアンケートも同様，Yahoo! クラウドソーシング[21]上で動物園や水族館に行ったことがある人を対象として，1,000 人に対して行ったアンケートである．なお有効回答は 989 件（男性 723 名，女性 257 名，無回答 9 名）であった．

アンケートの自由記述設問の内容と設問ごとの不真面目回答率を表2に示す．アンケートは全11問で構成され，内自由記述設問は 4 問であった．前節のアンケートと同様設問内容は，対象である動物園や水族館に行ったことがある人が必ず回答できるように構成されており，「特になし」といった回答ができないようになっていた．

不真面目回答率については，不真面目回答分類を行っていないため，前節と同様の基準で，著者が不真面目回答分類を行った．

4. 不真面目回答分類手法

本研究では ChatGPT を利用した 3 つの手法を用いて不真面目回答の自動分類を試みた．なお，それぞれ OpenAI の API[23]を利用して行った．利用したモデルは gpt-4-1106-preview である．

単純判定手法は，設問と回答を入力として，その設問に対して回答が真面目か否かを判定するものである．具体的には，回答が不真面目回答の基準として利用した「設問に対して答えが伴っておらず，回答そのもので意味を捉えられないもの」に当てはまる場合は不真面目回答として 1 を，そうでない場合は 0 と出力するように，回答が入力されると判定するようにした．なお実際に使用したプロンプトと出力フォーマットの指定文は表 3 の通りであり，『設問』にはアンケート上での設問の内容が入る．なお，出力結果があまり変わらないようにするため，モデルの temperature は 0 に設定した．

自信度指標手法は，単純判定手法と同様の不真面目回答の判定を出力させると同時に，その判定に対する自信度も同時に出力させるものである．自信度を同時に出力させることにより，分類の信頼性について評価することが可能となる．また Schonlau ら[19]が提案している半自動分類への応用が期待できる．実際に使用したプロンプトと出力フ

Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

表 3 手法ごとのプロンプトと出力フォーマットの指定文

手法番号	プロンプト	出力フォーマットの指定文
単純判定手法	あなたにはこれからアンケートで入力された回答が，『設問』という質問に対する回答として成り立っているかを回答してもらいます．私の発言に対して，以下のフォーマットで 1 回に 1 つずつ回答します．説明は書かないでください	質問に対して答えが伴っておらず，回答そのもので意味をとらえられないもの」に対しては 0 を，そうでないものは 1 と出力してください．
自信度指標手法		【判定】:「質問に対して答えが伴っておらず，回答そのもので意味をとらえられないもの」に対しては 1 を，そうでないものは 0 と出力してください． 【自信度】:判定の出力についての自信度を 5 段階（1 低い～5 高い）で出力してください．説明は不要です．
点数付与手法	あなたはインタビュアーとして，以下の指示に従ってください．あなたは『設問』という質問を，色々な人に聞いています．その質問に対する回答が入力されるので，その回答が返ってきて嬉しいかどうかを 100 点満点で出力してください．なるべく 100 点を出さず，厳しめに評価してください．出力は点数のみとし，理由を述べる必要はありません．	出力フォーマットの指定文はなし

ォーマットの指定文を表 3 に示す．なお自信度の出力の幅を増やすため，モデルの temperature は 0.5 と設定した．

点数付与手法は，回答に対して点数を付与するものである．インタビュアーという設定することで ChatGPT[11]が点数をつけやすいようした．また，具体的な状況と点数を嬉しさという基準で点数をつけさせた．理由としては，プロンプトを調整する際「質問に対して答えが伴っておらず，回答そのもので意味をとらえられないもの」という基準で点数をつけさせると，点数にばらつきがみられなかったこと，また点数を付与することにより，不真面目回答の閾値を設定することができるようになるためである．ここで，予備検討では点数にばらつきがみられなかったため，点数についても厳しめに評価するよう指示した．実際のプロンプトを表 3 に示す．なお，点数付与手法においては出力フォーマットの指定はしなかった．また，点数の基準があまり変わらないよう，temperature は 0 とした．

5. 評価実験

5.1 実験概要

ChatGPT[11]を用いた不真面目回答自動分類ができるのか，また効果的な自動分類方法について， 2 種類のアンケートデータを利用して，3 つの手法の判定精度の比較を行った．

本研究では評価指標として，正答率・再現率・適合率・F 値を用いた．ここで，正答率は「入力回答が真面目か不真面目かを正しく判定できた割合」，再現率は「正解ラベルが不真面目であるもののうち，正しく不真面目と判定できた割合」，適合率は「不真面目回答と判定したもののうち，正解ラベルが不真面目である割合」を表す．不真面目回答分類においては，再現率と適合率のバランスをとることが重要であることから，本研究では F 値を重要視する．

5.2 実験結果

まず手法ごとおよび，アンケートデータごとに最も精度が高かった条件の判定結果を表 4，5 に示す．表より，いずれのアンケートにおいても，正答率および F 値は点数付与手法，自信度指標手法，単純判定手法の順で高い結果となっていた．

単純判定手法を用いて，アンケートの設問ごとに判定を行った結果を表 6，表 7 に示す．表 6 より，F 値が最も高いのは Q-4 であり，再現率，適合率のバランスが取れている．一方で Q-1 は F 値が最も低く，特に適合率が低く，多くの真面目回答を不真面目回答と判断している．また表

表 4 運転免許を所持している人向けの
アンケートの手法ごとの判定結果

	単純 判定手法	自信度 指標手法	点数 付与手法
正答率	0.88	0.89	0.95
再現率	0.88	0.94	0.87
適合率	0.30	0.34	0.52
F 値	0.44	0.50	0.65

表 5 動物園・水族館に関するアンケートの
手法ごとの判定結果

	単純 判定手法	自信度 指標手法	点数 付与手法
正答率	0.83	0.87	0.93
再現率	0.83	0.77	0.63
適合率	0.22	0.28	0.41
F 値	0.35	0.41	0.50

Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

表 6 運転免許を所持している人向けの
アンケートの判定結果（単純判定手法）

	Q-1	Q-2	Q-3	Q-4
正答率	0.89	0.96	0.90	0.94
再現率	0.82	0.79	0.91	0.89
適合率	0.19	0.45	0.44	0.46
F 値	0.31	0.57	0.59	0.61

表 7 動物園・水族館に関するアンケートの
判定結果（単純判定手法）

	Q-1	Q-2	Q-3	Q-4
正答率	0.98	0.97	0.83	0.56
再現率	0.70	0.75	0.96	0.77
適合率	0.53	0.37	0.21	0.19
F 値	0.60	0.49	0.34	0.31

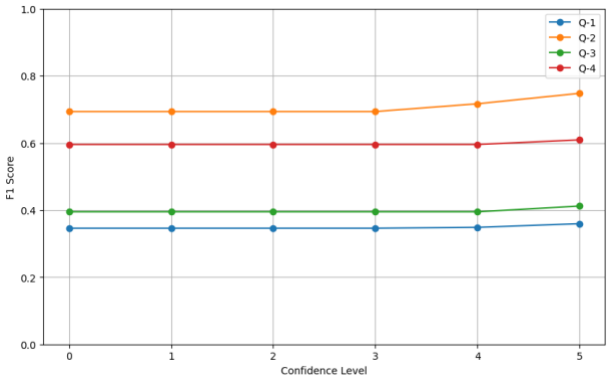

図 1 運転免許を所持している人向けの
アンケートの判定結果（自信度指標手法）

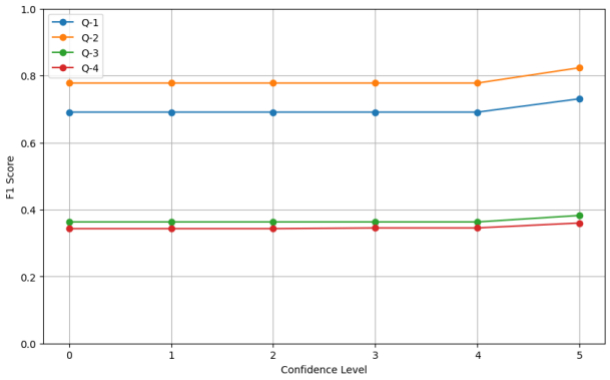

図 2 動物園・水族館に関するアンケートの
判定結果（自信度指標手法）

7 より，F 値が最も高いのは Q-1 である．一方で，Q-4 は F 値が最も低く，また正答率や適合率も他の設問と比べ低くなっている．

自信度指標手法を用いて，アンケートの設問ごとに判定を行った結果を図 1，図 2 に示す．グラフの横軸は自信度の閾値，縦軸は F 値を示している．図より，どちらのデータセットにおいても，自信度が 5 の場合の方が 4 と 3 に比べ F 値の値が高くなっている．また設問ごとに見ると，どちらのデータセットにおいても Q-2 の F 値の値が最も高い結果となっていた．しかし，あまり大きな差がないこともわかる．

点数付与手法を用いて，アンケートの設問ごとに判定を行った結果を図 3，図 4 に示す．グラフの横軸は点数の閾値，縦軸は F 値である．なお，F 値は点数の閾値以上のものを真面目回答，未満のものを不真面目回答とし，F 値を算出した．図 3 より，Q-1，Q-4 の F 値は閾値が 10 点の時に最も高い結果となっていた．一方，Q-2，Q-3 は閾値が高くなるにつれて，F 値が減少していた．図 4 より Q-1，Q-3 の F 値は閾値が 10 点の時に，Q-2 は 30 点の時に最も高い結果となっていた．一方，Q-4 は 50 点の時に F 値の値が上昇するが，閾値が 0 点の時が最も F 値が高い結果となっていた．

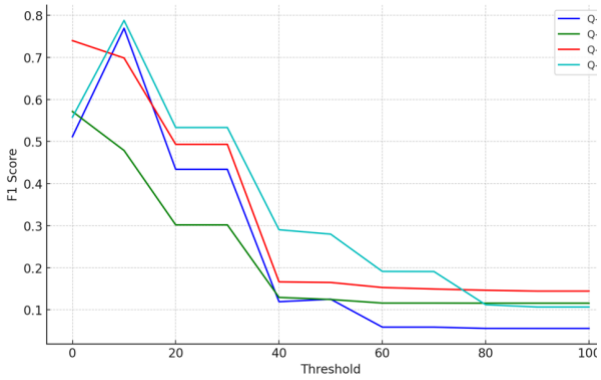

図 3 運転免許を所持している人向けの
アンケートの判定結果（点数付与法）

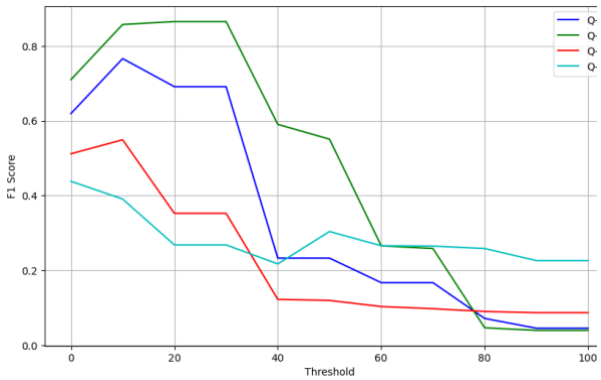

図 4 動物園・水族館に関するアンケートの
判定結果（点数付与法）

Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

6. 考察

6.1 総合的な結果について

結果より，正答率および F 値は点数付与手法，自信度指標手法，単純判定手法の順で高い結果となっていた．このことから，ChatGPT を用いて不真面目回答分類を行う場合は，単純に判定させるだけではなく，工夫が必要であると考えられる．また，F 値の値は全体的に低い結果であったため，不真面目回答分類の指標として用いる場合には，さらなる工夫が必要であると考えられる．

また設問ごとの F 値については設問ごとに大きく異なった．その理由として，設問ごとの特性が影響していると考えられる．例えば，動物園・水族館に関するアンケートにおける Q-1 は，回答者が単語ベースで回答する設問であり，F 値は他と比べ高かった．一方で，Q-3 や Q-4 など文章ベースで回答する設問では F 値が低かった．本来は文章ベースでも精度よく判定できることを期待していた．そのため今後は文章ベースでも精度よく判定できるように，判定手法の検討を行っていく．

6.2 手法間比較

結果より，単純判定手法，自信度指標手法を比較すると，F 値が大きく異なっていた．手法に着目すると，自信度の判定を除き，temperature パラメータの設定が単純判定手法は 0 に対し，自信度指標手法は 0.5 と異なっている．このことから，temperature の設定が F 値に大きく影響することが考えられる．

また結果より，点数付与手法の閾値 10 点の F 値の値が，他の手法と比べ高い結果となっていた．特に，閾値 0 の時に比べ，F 値の値が大きくなっていた設問に着目すると，その他の設問と比べ不真面目回答率が低い設問であったことがわかる．このことから，設問の形式にもよるが，不真面目回答がそもそも集まらないような設問や，不真面目回答が集まりにくいアンケートにおいて，この判定手法が有効である可能性がある．

6.3 判定結果について

結果より，ChatGPT による分類精度は期待したほど良いものではなかった．その理由として，今回利用したアンケートがそもそも不真面目回答率が低く，一つの判定ミスが結果に大きく影響したことが考えられる．例えば，動物園・水族館に関するアンケートの Q-1 の設問において「とうぶどうぶつこうえん」，「しんえのしますいぞくかん」といった，本来漢字表記があるものをひらがなで回答しているものが誤って不真面目回答と判定されてしまっていた．これは，ひらがなによる回答が，子供っぽさを演出するものになっており，そのことが不真面目であると判断された可能性がある．こうした表記の扱いについては今後検討を行っていく必要がある．

一方，「特になし」といった回答をできないような設問に限定していたが，実際は「特になし」の回答が含まれており，このような回答も分類にかけたところ，分類結果は不真面目回答とうまく判定することができていた．また，「上野」，「品川」といった動物園・水族館を表記せず地名だけの回答に対して，正解ラベルが真面目回答であり判定結果も真面目回答であった．このような回答は，分類者の事前知識が必要となる回答であり，こうした回答に対しての分類は ChatGPT の利用が有効である可能性が考えられる．

また結果より，動物園・水族館に関するアンケートの Q-4 では，正答率が低い結果となっていた．これは，リピート訪問をした，もしくはしなかった理由について回答を求めていた点について，リピート訪問をした回答者の多くが施設名のみを回答しており，設問を満たしていないため，正解ラベルを不真面目としていた．しかし，ChatGPT を用いた判定では，それらの多くが真面目回答に分類されており，理由を回答しなければならないことを十分に把握できていなかったものと考えられる．その結果，他の設問よりも精度が良くなかったと考えられる．そのため，こうした回答を完全に満たしていない回答に対しては，ChatGPT に対し，何をどう答えることが望まれているのかについて判断させることが重要であると考えられる．

7. まとめ

本研究では，Web アンケートなどにおける自由記述設問の不真面目回答を，ChatGPT を用いて自動分類することが可能であるかを調査することを目的として研究を行った．具体的には，単純に不真面目判定する手法，自信度を同時に出力する手法，点数をつけ判定する手法を提案し，ChatGPT を用いた不真面目回答自動分類ができるのか検証を行った．実験の結果，回答に対して点数を付与する点数付与手法を利用し，10 点より高いものを真面目，低いものを不真面目回答とする手法が最も分類精度が高いことがわかった．一方，F 値は十分ではなく，判定における課題が残った．

今後は，より判定精度を上げるため，プロンプトの調整や，新たな判定手法について再検討していく予定である．また，より不真面目回答率の高いアンケートデータに対して本手法が有効かについても検証していく予定である．さらに，単純に人手でアノテーション付けをするのではなく，ひとと ChatGPT が連携して合議によりアノテーション付けを行っていくような手法も検討予定である．

Vol.2024-CN-121 No.22
Vol.2024-CDS-39 No.22
Vol.2024-DCC-36 No.22
2024/1/22

参考文献

[1] Vergnaud, A. C., Touvier, M., Méjean, C., Kesse-Guyot, E., Pollet, C., Malon, A., Castetbon, K. and Hercberg, S. “Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Santé study.” International Journal of Public Health, Vol. 56, No. 4, p. 507-417 (2011).

[2] Reja, U., Manfreda, K., Hlebec, V., Vehovar, V. “Open-ended vs. Close-ended Questions in Web Questionnaires.” Adv Methodol Stats, Vol. 19, No. 1, p. 159-177 (2003).

[3] Schuman, H., Presser, S. “The Open and Closed Question.” American Sociological Review, Vol. 44, No. 5, p. 692-712 (1979).

[4] Couper, M. P., Kreuter, F. “Using paradata to explore item level response times in surveys.” Journal of the Royal Statistical Society, Vol.176, No. 1, p. 271-286 (2013).

[5] Yan T., Tourangeau R. “Fast times and easy questions: The effects of age, experience and question complexity on Web survey response times.” Applied Cognitive Psychology, Vol. 22, No. 1, p. 51-68 (2008).

[6] Revilla, M., Ochoa, C. “Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role?” Quality & Quantity, Vol. 50, No. 6, p. 2495-2513 (2016).

[7] Holland, J. L., Christian, L. M. “The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys.” Social Science Computer Review, Vol. 27, No. 2, p. 196-212 (2009).

[8] Zuell, C., Menold, N., Körber, S. “The Influence of the Answer Box Size on Item Nonresponse to Open-Ended Questions in a Web Survey.” Social Science Computer Review, Vol. 33, No. 1, p. 115-122 (2015).

[9] Yamazaki, I., Hatanaka, K., Nakamura, S. and Komatsu, T. “A Basic Study to Prevent Non-Earnest Responses in Web Surveys by Arranging the Order of Open-ended Questions.” International Conference on Human-Computer Interaction (HCII 2023), LNCS, Vol. 14011, p. 314-326.

[10] Richard M. Durand, Zarrel V. Lambert. “Don’t know responses in surveys: Analyses and interpretational consequences.” Journal of Business Research, Vol. 16, No. 2, p. 169-188 (1988).

[11] ChatGPT. 入手先〈https://chat.openai.com/〉, (参照: 2023-12-20).

[12] Ronggang, Z., Xiaorui, W., Leyuan, Z., Haiyan, G. “Who tends to answer open-ended questions in an e-service survey? The contribution of closed-ended answers.” Behaviour & Information Technology, Vol. 36, No. 12, p. 1274-1284 (2017).

[13] Schmidt, K., Gummer, T., Roßmann, J. “Effects of Respondent and Survey Characteristics on the Response Quality of an Open-Ended Attitude Question in Web Surveys.” Methods, Data, Analyse, Vol. 14, No. 1, p. 3-34 (2020).

[14] Galesic, M., Bošnjak, M. “Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey.” Public Opinion Quarterly, Vol. 73, p. 349-360 (2009).

[15] Smith, T. W. “Little Things Matter: A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses.” In Proceedings of the American Statistical Association, Section on Survey Research Methods, p. 1046-1051 (1995).

[16] Emde, M., Fuchs, M. “Using Adaptive Questionnaire Design in Open-Ended Questions: A Field Experiment.” Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research (AAPOR), 2012.

[17] 山﨑 郁未, 畑中 健壱, 中村 聡史, 小松 孝徳. “自由記述設問の順番とテキストボックスサイズが離脱に及ぼす影響:スマートフォン・PC の比較.” 研究報告ヒューマンコンピュータインタラクション（HCI）, Vol. 2023-HCI-204, No. 18, p. 1-8 (2023).

[18] Kawamoto, T., Aoki, T. “Democratic classification of free-format survey responses with a network-based framework.” Nature Machine Intelligence, Vol. 1, No. 7, p. 322–327 (2019).

[19] Schonlau, M., Couper, M. P. “Semi-automated categorization of open-ended questions.” Survey Research Methods, Vol. 10, No. 2, p. 143-152 (2016).

[20] Gweon, M., Schonlau, M. “Automated Classification for Open-Ended Questions with BERT.” Journal of Survey Statistics and Methodology, 2023.

[21] “Yahoo!クラウドソーシング“. 入手先〈http://crowdsourcing.yahoo.co.jp/〉, (参照 2023-12-20)

[22] 畑中 健壱, 山﨑 郁未, 中村 聡史. “ShrinkTextbox: Web アンケートの自由記述回答欄サイズ変化による回答の質向上法.” 研究報告ヒューマンコンピュータインタラクション（HCI）, Vol. 2023-HCI-201, No. 20, p. 1-8 (2023).

[23] “OpenAI API”. 入手先〈https://openai.com/blog/openai-api〉, (参照 2023-12-20)