

ChatGPTを用いたメディアの内容分析



早稲田大学
政治経済学部准教授 中村 理

1. はじめに

筆者の担当する学部4年の専門ゼミ演習では、メディアに用いる内容分析という手法をChatGPTに委ねる取り組みをおこなっています。それを以下に紹介します。

2. 内容分析とは

内容分析とは、メディアに流れる情報をコード(記号)化して集計し、分析するものです。例えば、新聞記事に登場する人物を「政治家」や「専門家」といったコードに分類したり、SNSにみられる論調を「ポジティブ」や「ネガティブ」といったコードに分類したり、です。コードを集計することで、情報の特徴を見出すことができます。コードに分類する作業をコーディングといいます。

内容分析には主に2つの技法があります。第1はヒューマン・コーディング、すなわち人がコーディングをするものです。第2はコンピュータ・コーディング、すなわちコンピュータでコーディングをするものです。それぞれにはメリットとデメリットがあります。

第1のヒューマン・コーディングのメリットは、細かな分析ができる点です。例えば上にあげた登場人物や論調、ほかにも争点など、研究の目的に応じて必要なコード化をおこなうことができます。分析対象もテキストに限りません。雑誌の表紙で著名人がとるポーズや表情もコード化できます。一方、デメリットは、たくさんの分析対象を扱うことはできない点です。コーディングを人手でおこなうため、数には限界があります。

第2のコンピュータ・コーディングのメリットとデメリットは、その逆です。メリットは、大量のテキストを短時間で客観的に分析できる点です。一方、デメリットは、これまでのところは細かな分析ができない点です。特に、意味的な判断をとるコーディングがコンピュータは苦手です。例えば、単純に「新型コロナウイルス」という語が登場した回数を数えることは得意です。けれども、新型コロナウイルスの広まりの現状に言及されたのか、対策に言及されたのか、といった争点を分類するとなると精度が落ちるのです。ど

の争点でポジティブだったかを判断することも得意ではありません。人の判断にもゆらぎがありますが、コンピュータで安価に実現できるコーディングにはより注意が必要というのが現状です。

3. ChatGPTによる内容分析

ChatGPTをはじめとする生成AIは、上にあげるこれまでのコンピュータ・コーディングのデメリットを解消できる可能性を持ちます。そこで、筆者の2024年度学部4年の専門ゼミ演習で、以下のように学生にChatGPTによる内容分析を試してもらいました。受講者は16名、プログラミングに関する知識は多くの学生にない状態、全体にかかった授業時間は協調型学習のもとで約90分でした。

今回の直接の授業目標は、「大阪万博に関するX(旧Twitter)上のポストが万博開催にポジティブかネガティブかをChatGPTに判断してもらおう」とこととしました。この狙いには、ChatGPTを分析に用いる方法を体験することで、興味を持った学生に今後の卒業研究で役立ててもらうことがあります。

学生にはまず、Xから大阪万博に関するポストを集めてもらいました。今回は試行のため、一定の条件下で10から20ほどという、ごく少数のポストを手作業で得てもらいました。Xを試行対象にした理由は4つあります。第1は短文のため学生が内容を視認しやすいこと、第2は同じく短文のためChatGPTで処理するコストを低くできること、第3は学生が興味を持ちやすいSNSの1つであること、第4は教室で容易にアクセスできることからです。今後、ChatGPTを研究に用いる場合は、対象をXに限る必要はなく、数も増やせば良いでしょう。

次に、学生にはChatGPTの開発元であるOpenAIにアカウントを設定してもらいました。また、OpenAIのAPIというものを利用するためのKeyを得てもらいました。APIとは、ChatGPTに外からアクセスするための窓口にあたるものです。アカウント開設については、インターネット上でガイドをしてくれているサイトが複数あります。学生には、利用にかかる料金の確認、上限額

の設定も、同時に済ませてもらいました。

次に、学生にはGoogle Sheetsへ行き、OpenAIのAPIを経由してChatGPTにアクセスするための関数を作ってもらいました。この方法についても、インターネット上でガイドをしてくれているサイトが複数あります。基本的にはそうしたガイドが例示する関数を、自分のGoogle Sheets上に持ってくる作業になります。上で得たKeyはここで用います。

次に、学生には、さきほど得たXのポストをGoogle Sheetsに貼り付けてもらいました。

最後に、学生にはそれらポストの一つ一つを、「次の文書は大阪万博の開催に対してポジティブか、ネガティブか、いずれでもないか、判別してください」という指示文とともに、ChatGPTへ投げてもらいました。ChatGPTへ投げる作業は、上で作ったGoogle Sheets上の関数に、この指示文とポストを入力することでおこないます。そうすると、ChatGPTからの答えが「ポジティブ」などとGoogle Sheetsに戻ってきます。Google Sheets上では、先頭のポストでこの作業をおこなった後、オートフィルをすることで、続くすべてのポストの結果を短時間のうちに得ることができます(表1に例示)。今回は少数のXのポストだけで試しました。しかし、数千～数万のポストや文書でも時間はかかりません。

精度も良好です。筆者とChatGPT(GPT-4)のそれぞれが大阪万博に関するX上のポスト50個を読み、誘致にポジティブかネガティブかを判断しま

した。その結果、両者の一致度は、クリッペンドルフの α という指標で0.75でした。一般に、人同士でポジティブ・ネガティブを比較したときの α が0.7から0.8程度です。したがって、これは研究に耐える数値と言えます。ChatGPTは「誘致に」という指定も理解していました。

4. ChatGPTを内容分析に用いるメリットと留意点

このようにChatGPTを内容分析に用いるメリットは3つあります。第1は、メディアの内容分析において、これまでコンピュータ・コーディングが苦手としてきた意味的解釈を要する部分に手が届きつつある点です。今回は大阪万博という対象を指定してポジティブかネガティブかを判断してもらいました。同様に、争点を判断してもらったり、判断の理由を述べてもらったりすることもできます。

第2は、応用の幅が広い点です。指示文を変えただけで、様々なコーディングを実現できる余地があります。複数の言語を同時に扱うこともできます。また、将来的にはテキストだけでなく、画像や映像も分析対象として安価に扱えるようになるでしょう。

第3は、利用者である学生に難しいスキルが求められない点です。Excel(やGoogle Sheets)を使える程度のスキルがあれば、1時限のうちに分析を終えるところまでを体験できます。プログラミングの素養は必須ではありません。こうした体験は、学生が将来、研究やビジネスに進む局面で役に立つはずで

一方で、教室でこの学びを導入するには2つ、留意点もあります。第1は、OpenAIのAPIを利用するにあたり、学生はクレジットカード等で課金する必要がある点です。参考までに、ここで紹介した内容(10～20ポストのみの分析)の場合、gpt-4o(現時点で高額側のChatGPTのモデル)を用いた課金総額は0.1～0.2ドル程度です。額は大きくないものの、一般の教室で学生に課金登録をしてもらうハードルは高いものと想像します。

第2は、ChatGPTの解答にいたる過程はブラックボックスである点です。そのため、研究として用いる場合は、なんらかの方法で結果の信頼性と再現性に言及する必要があります。

5. 終わりに

以上、課題はあるものの、学生が研究を進めるにあたってChatGPTをはじめとする生成AIは内容分析に大きな果実をもたらすことが期待されます。皆さまとの情報共有がより効果的・効率的な授業運営と研究の進展につながれば幸いです。

表1 XのポストとChatGPTの判断例

| ポストの内容 | 判断 |
|--|---------|
| 「大阪もいつ地震が起きるか分からない状況なのに、これで万博なんて本気で開催する気なのかしら」とツイートすると、一部の方々の逆鱗に触れるらしい。不謹慎な人間で悪うございましたね。 | ネガティブ |
| #エコノモーニング 金曜日は経済ジャーナリスト内田裕子さんの経済情報！ 今朝は「2025年万国博覧会 大阪の誘致活動」についてでした。実現すれば20年ぶりの万博！時代の先を行く新しい形の万博が開催されるかも！みんなで盛り上げて、実現させたいですね！ #nack5 #レディオファントム | ポジティブ |
| 大阪の(市営)地下鉄では半世紀前頃まで黙認されていたそうですが、万博開催前に取り締まったのだと聞いたことがあります。まあそうは言いながら、関西のどこかの路線の某駅では、公式(違)の自販機に隣接して回数券を一枚ずつ売る金券ショップの自販機が設置されていたりもするようですけれど(笑) | いずれでもない |