

COVID-19 流行下における不満の時期別特徴分析

小川和浩, 中塚晶仁, 荒川健, 池田貴康, 河野吉宏, 富樫健太, 浪岡保男 (東京都立産業技術大学院大学)

背景・目的

有事下の世相の変遷を速やかに捉えることは有意義だが、実施には多くの時間とコストが掛かる。

- ◆ 新型コロナ不満アンケートデータのような自由記述のアンケートから、回答内容の傾向を定量的に把握するには記述内容が類似の回答をまとめ、グループごとの数を集計するアフターコーディングが必要。しかし、複数の意見の混在、文章の曖昧性の高さなどの要因があり単純な作業ではない。そのため、人の手でアフターコーディングを行うと回答数に比例して多くの時間とコストが掛かる。
- ◆ アンケート結果を分析する際、多数意見・小数意見のどちらも把握することが重要
 - 多数意見を把握する目的：全体の傾向や多数の人々が抱える共通の課題を明らかにする。
 - 少数意見を把握する目的：特定の層やニッチな課題を明らかにする。

◆ 複数の意見が含まれている回答の例

満員電車で当たる時間帯に、出勤したくないので、出勤時間を遅らせて欲しい。また、帰宅時間も、満員電車で被らない時間に変更してほしい。コロナが怖いので、オンラインで、自宅で仕事できるようにするか、選べるようにしてほしい。

出勤と帰宅の時間を電車が混雑する時間帯からずらしたい

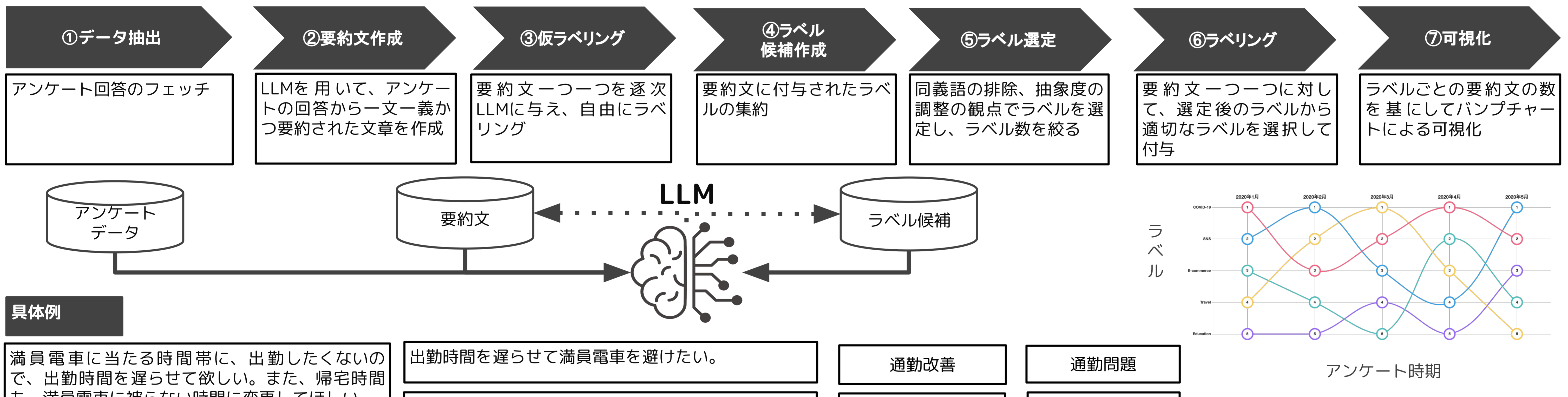
リモートワークをしたい。

- ◆ 大規模言語モデル (LLM) の進展によって、作業に時間とコストが掛かる2つ作業の効率化が図れるようになった。
 - 文意を保持した文章の要約
 - 分類や集計のためのラベルづけ

LLMを用いたアフターコーディングとアンケート内容の傾向把握のための可視化手法の検討

提案手法

- ◆ LLMの用いた文章要約、ラベルリングによる文脈情報の作成と可視化の半自動化。

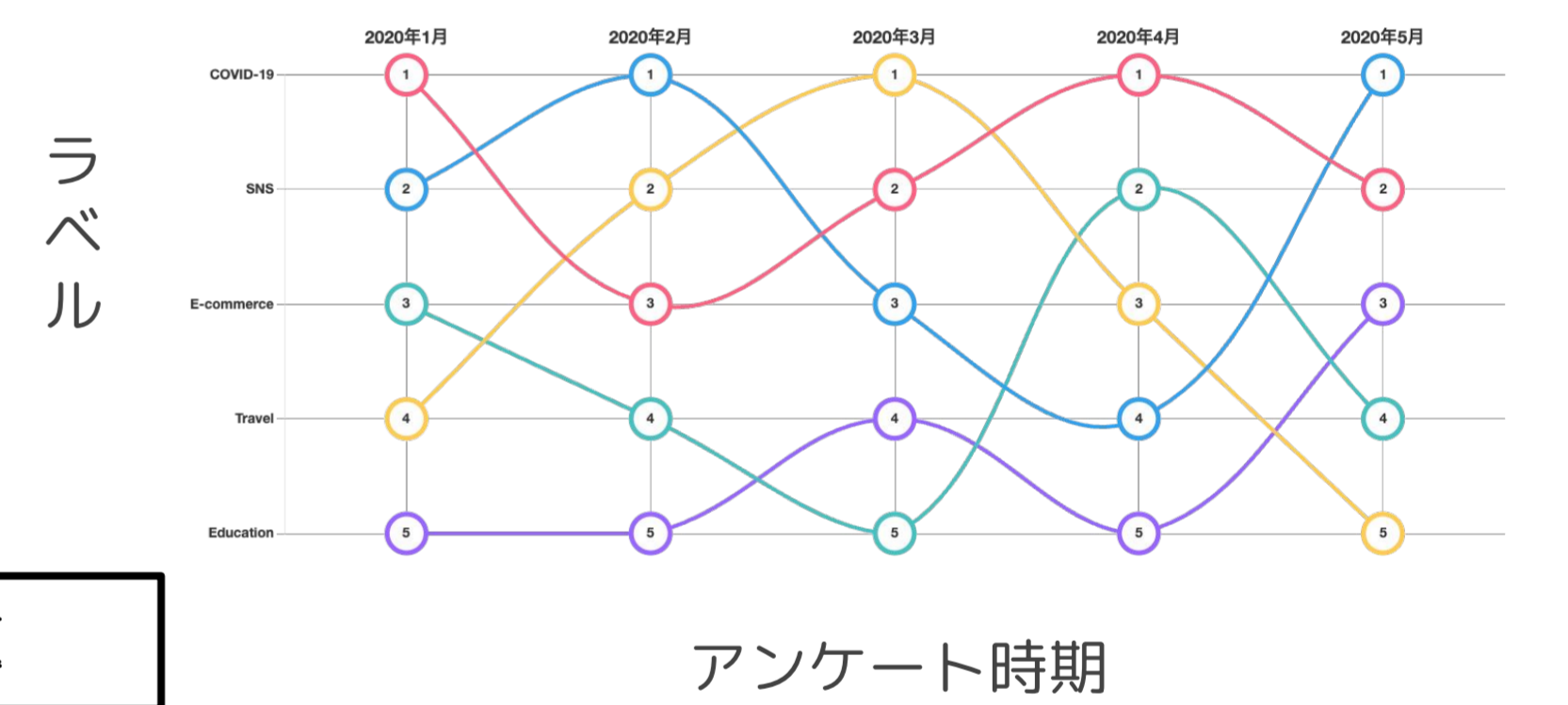


具体例

満員電車で当たる時間帯に、出勤したくないので、出勤時間を遅らせて欲しい。また、帰宅時間も、満員電車で被らない時間に変更してほしい。コロナが怖いので、オンラインで、自宅で仕事できるようにするか、選べるようにしてほしい。

出勤時間を遅らせて満員電車を避けたい。
帰宅時間を変更して満員電車を避けたい。
オンラインで自宅勤務を選べるようにしてほしい。

通勤改善	通勤問題
通勤改善	通勤問題
在宅勤務	リモート希望

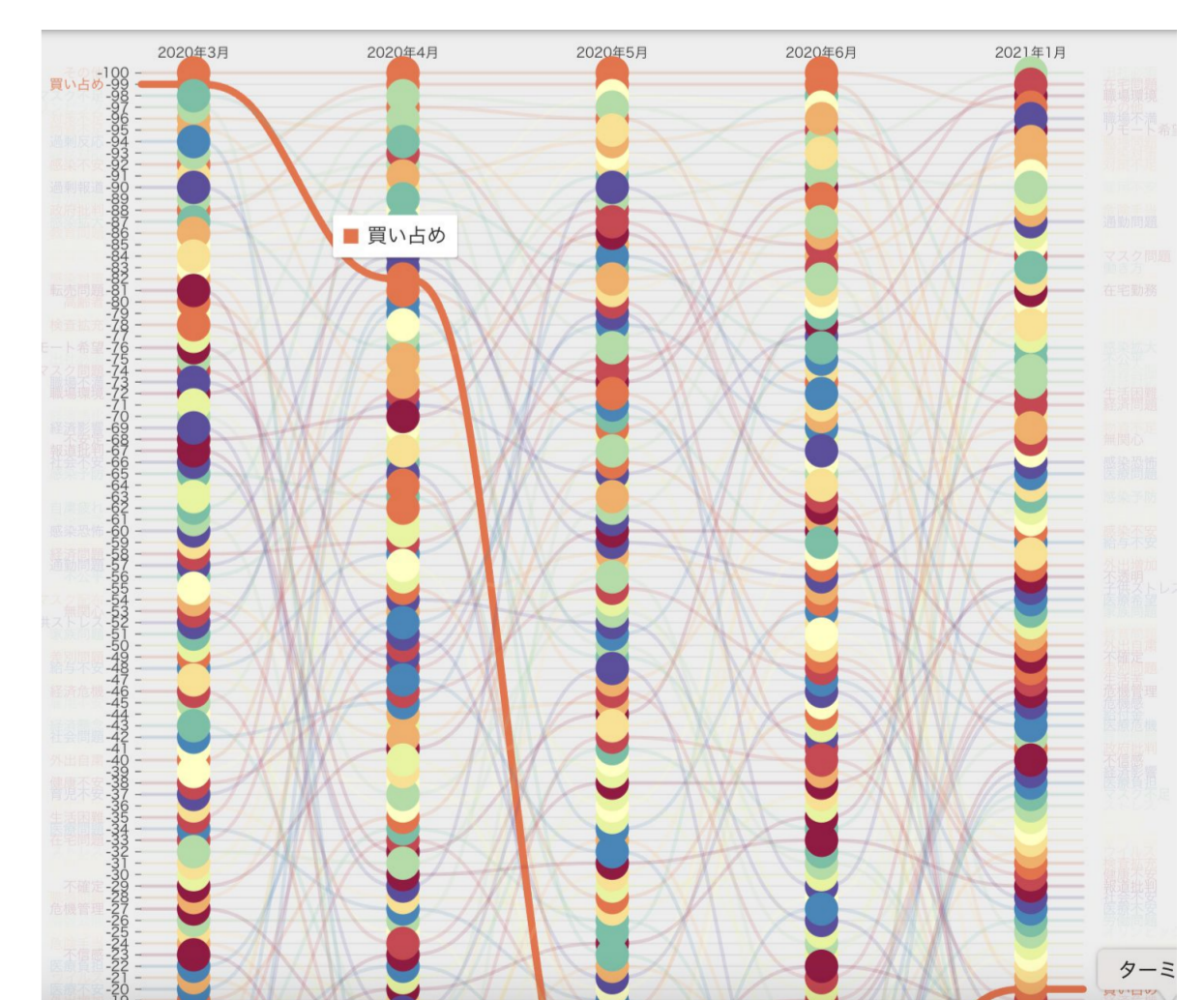


評価とその結果

- ◆ データセット
新型コロナ不満アンケートデータ (prefectureが「東京都」で、answerがNULLのデータを除外した2005件を分析対象とした)。
 - ◆ GPT-4o (gpt-4o-2024-11-20) を使用。
 - ◆ 提案手法の要約文作成がどの程度適切に行われたかを検証するために、意見分析に適した意見タグ獲得改善への取り組み[1]を参考に、日本語流暢性 (文章が日本語として自然かどうか) と情報保持性 (文章が原文の情報を十分に保持しているかどうか) を確認し、ともに高精度であることを確認。
- | 検証項目 | 精度 | 件数 |
|--------|-------|---------|
| 日本語流暢性 | 0.997 | 350/351 |
| 情報保持性 | 0.838 | 294/351 |
- 精度 = (正しく要約されている文章の数) / 検証した文章の数
- ◆ 最終的に用いられたラベルの数は100 (「その他」を含む)
 - “その他”に分類された文章は397件 (7.76%)
 - 適切なラベルが付与されていた割合は94.59%(332/351)

考察・展望・課題

- ◆ ラベルによる意見の推移が捉えられているかを検証するため「買い占め」のアンケート時期による推移を確認。
 - 政府による転売対策が実施された時期に合わせて意見の数が下落していることが読み取れる。
 - アンケート時期の政府の対応
 - 2020年3月マスク転売禁止
 - 2020年5月アルコール消毒製品の転売禁止
- ◆ 展望・課題
 - ラベリング機能の充実 (重複の排除、適切なラベル数算出)
 - UIの拡充 (ラベル→要約文→原文へのアクセス、LLMを活用した要因分析)
 - LLMへのプロンプトの深耕、トークンの削減
 - 情報保持性の精度向上



参考文献

[1]三澤賢祐：言語処理学会 第24回年次大会 発表論文集, pp.572-575 (2018.03).
 [2]Insight Tech Lab (インサイトテックラボ). 新型コロナウイルスに関する調査レポート. <https://lab.insight-tech.co.jp/articles/192/> (2020.03).
 [3]Insight Tech Lab (インサイトテックラボ). 第2回新型コロナウイルスに関する調査レポート. <https://lab.insight-tech.co.jp/articles/197/> (2020.04).
 [4]Insight Tech Lab (インサイトテックラボ). 第3回新型コロナウイルスに関する調査レポート. <https://lab.insight-tech.co.jp/articles/216/> (2020.05).
 [5]Insight Tech Lab (インサイトテックラボ). 第4回新型コロナウイルスに関する調査レポート. <https://lab.insight-tech.co.jp/articles/221/> (2020.06).
 [6]Insight Tech Lab (インサイトテックラボ). 第5回「新型コロナウイルス」に関する意識調査. <https://lab.insight-tech.co.jp/articles/250/> (2021.01).

謝辞

「本研究では、国立情報学研究所の情報学研究データリポジトリ (IDR)によるデータセット提供サービスを通じて、株式会社Insight Tech様から提供を受けた「不満調査データセット」 (<https://www.nii.ac.jp/dsc/idr/fuman/>)を利用させていただきました。心より感謝いたします。