

# COVID-19不満調査に関する自由記述アンケートの自動集約

P12 銭本友樹, 長谷川遼, 宇津呂武仁 (筑波大学)

## 背景・概要・データセット

**背景:** 自由記述アンケートは、調査対象における新たな価値観や意見の発見に貢献する重要な手法である。その分析は一般に人手で行われるが、大規模な回答の分析には莫大な時間と費用が必要である。

**概要:** 大規模言語モデルを用いて自由記述アンケートの回答を自動で分析する手法を提案した。人手を用いた分析手法と提案手法を性能・費用・時間の面で比較し、提案手法の有用性を示した。

**データセット:** 「不満買取センター」上で実施された、「新型コロナ不満アンケートデータ」(2020年3月: 2,996回答, 2020年6月: 2,997回答)を使用

## 質的研究 / グラウンデッド・セオリー・アプローチ [戈木クレイグヒル, 2014] に基づく分析

**グラウンデッド・セオリー・アプローチ:** データから概念を抽出し、概念同士を関連付けようとする方法

**対象質問:** 「新型コロナウイルス」に関して懸念していることや不満をお知らせください。

### ステップ1: データの切片化

データを異なる内容ごとに分割 (切片化) する。

### ステップ2: プロパティとディメンションの抽出

切片化データごとに、プロパティ (データ中の様々な概念) とディメンション (プロパティの実際の値) を抽出する。

### ステップ3: ラベルの付与

プロパティとディメンションをもとにデータの内容を端的に表すラベル名を付与する。

### ステップ4: カテゴリの作成と分類

類似したラベルをまとめ、それらを端的に表すカテゴリ名を付与する。

### 大規模自由回答アンケート

人手でデータ全体を読み込み、何度も手戻りをしながら慎重に各手順を行う

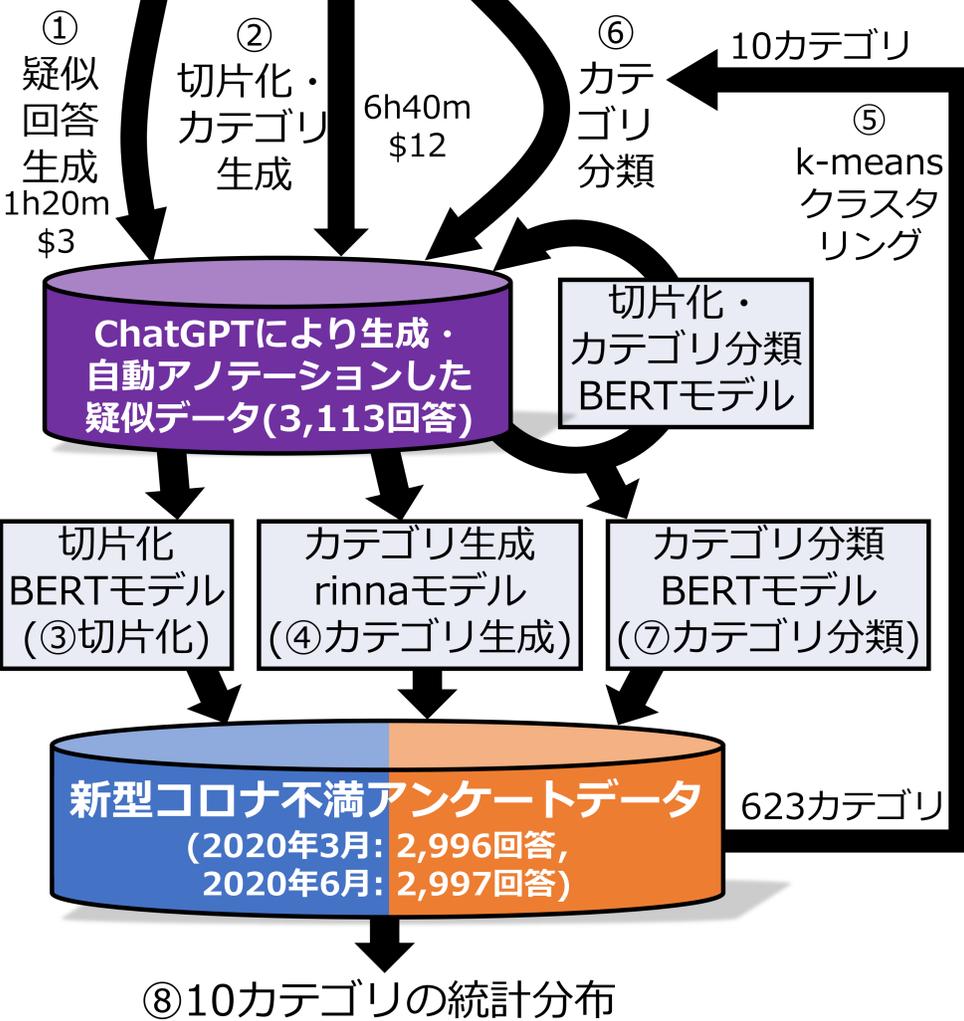


切片化データ	プロパティ: ディメンション	ラベル	カテゴリ
学校が休みになり、子供の勉強の事が心配...小学校の対応が悪い。	感情: 心配, 学校の状態: 休み 心配の対象: 子供の勉強, 小学校の対応: 悪い	小学校の対応への不満	教育
出かけるところがへった。	出かけるところ: 減った	外出の難しさへの不満	日常生活

## 自動分析モデルの構築と適用の流れ

### ChatGPT (gpt-3.5-turbo-0613)

人手(時給1,500円)と比べ、**時間:1/30, 費用:1/350**, でのアノテーションが可能



**貢献1:** 疑似データ訓練モデルは人手アノテーションデータ訓練モデル及びChatGPTに匹敵する性能を達成

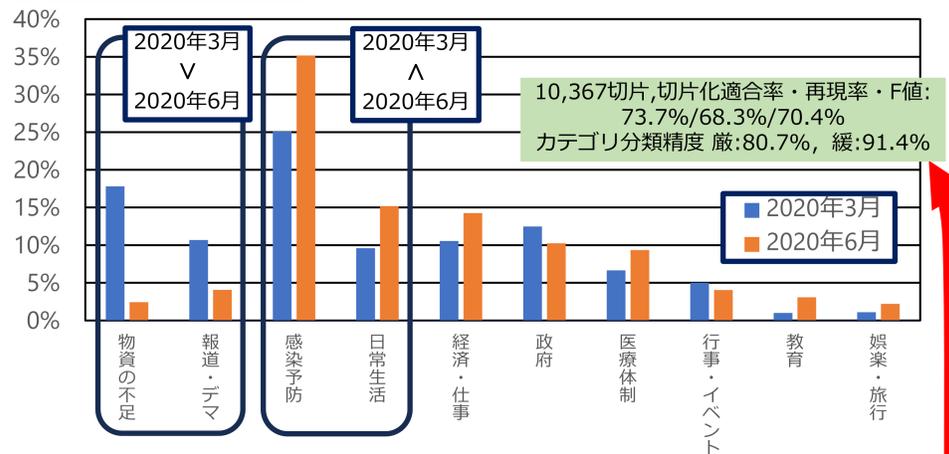
### ChatGPTにより生成した疑似回答に人手でアノテーションした評価データを対象とした性能比較

モデル	切片化適合率・再現率・F値	カテゴリ分類精度
人手アノテーションデータで訓練したBERTモデル	64.4%/66.1%/65.2%	※1 厳:83.6%, 緩:93.8%
疑似データで訓練したBERTモデル	38.9%/93.5%/54.9%	厳:76.9%, 緩:90.5%
ChatGPT	64.2%/57.0%/60.4%	厳:74.3%, 緩:88.2%

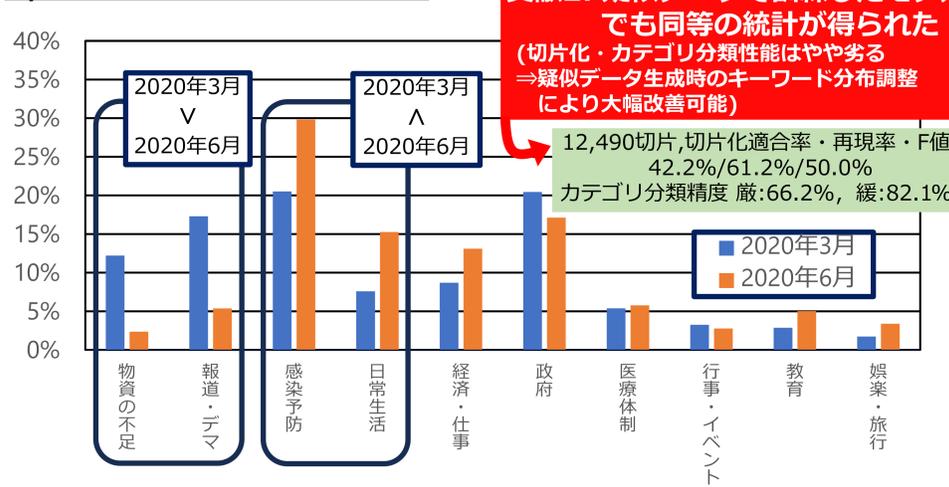
※1: 回答データには「デマのせいでマスクが全然買えない」のように、「物資の不足」と「報道・デマ」などの複数のカテゴリに当てはまる意見が多く存在する。そのため各回答には、その回答を代表とする1つのメインカテゴリと、その回答が当てはまる複数のカテゴリをアノテーションしている。「厳」では予測カテゴリとメインカテゴリとの一致を判定しており、「緩」では複数のカテゴリのどれかとの一致を判定している。

## 実データにおける統計分布の比較

### 人手でアノテーションした実データ832回答での訓練モデル



### ChatGPTにより生成・自動アノテーションした疑似データ3,113回答での訓練モデル



**貢献2:** 疑似データで訓練したモデルでも同等の統計が得られた (切片化・カテゴリ分類性能はやや劣る ⇒疑似データ生成時のキーワード分布調整により大幅改善可能)

## 結論・謝辞

**結論1:** ChatGPTは人に匹敵する精度でのアノテーションを非常に安価・短時間で実現。

**結論2:** ChatGPTにより生成・アノテーションした疑似データで訓練したモデルは、ChatGPTと同等の性能を実現。

**結論3:** 疑似データで訓練したモデルは、実データで訓練したモデルと同等の分析結果(カテゴリの統計)を実現。

**今後の展望:** プロパティ・ディメンション・ラベル生成の評価。

**謝辞:** 本研究では、NIIのIDRデータセット提供サービスにより株式会社Insight Techから提供を受けた「不満調査データセット」を利用した。