

[共通セッション] 土木分野におけるAIの活用

2024年9月6日(金) 10:40 ~ 12:00 第2講義室(経済学部)(川内南キャンパス文科系総合講義棟)

土木分野におけるAIの活用 (その7)

座長：植田 知孝 (オリエンタルコンサルタンツ)

11:40 ~ 11:50

[CS14-50] Claude3を用いたxROAD道路橋データの不一致箇所検知可能性の検証

*高橋 悠太¹ (1. 八千代エンジニアリング株式会社)

キーワード：道路橋デジタルツイン、大規模言語モデル、Claude3、xROAD

土木のアナログデータをデジタルに変換する際、転記ミス等によりフォーマット間で一致しないデータが生じる。デジタル化初期によく発生することから、この不一致データを発見し修正する手法を早期に見出すことが望ましい。先行研究では大規模言語モデルを用いたチャットボットであるChatGPTに注目し、xROAD道路橋データにおける不一致データの検知手法として提案した。2024年3月段階において、ChatGPT同様に日本語の読み取り生成、ファイルの読み取りが可能なAPIとしてClaudeシリーズが挙げられる。本研究ではClaude3を先行研究と同様の課題に適用した結果を分析考察することで、適用可能性を検証する。

Claude3 を用いた xROAD 道路橋データの不一致箇所検知可能性の検証

八千代エンジニアリング 正会員 ○高橋 悠太

1. 背景

Digital Twin は物理空間と仮想空間の双子を意味し、物理的に作用可能なリアル空間と演算可能なデジタル空間とをセンサー等で得られたデータで接続する¹⁾。この時、データはセンサー以外にも、人によって記入・集計される場合もある。センサーで得られたデータにも誤作動による異常データが含まれることがあるが、人によるデータも記入ミス等によって実際とは異なるデータが入力されうる。このように紙とデータ等、異なるフォーマット間で一致なくなってしまうデータはデジタル化初期によく発生することから、この不一致データを発見し修正する手法を早期に見出すことが望ましい。

先行研究²⁾では言語モデル大規模言語モデル(LLM: Large Language Model)を用いたチャットボットである ChatGPT³⁾に注目し、不一致データの検知手法として提案した。2024年3月段階において、ChatGPT 同様に日本語の読み取り生成、ファイルの読み取りが可能な API として Claude シリーズが挙げられる。本研究では Claude3⁴⁾を先行研究と同様の課題に適用した結果を分析考察することで、適用可能性を検証する。

2. 実験手順

道路橋に関するデータは HP 上で xROAD の道路橋データベース・橋梁リスト (77 条調査) より取得できる⁵⁾。同ページでの検索結果で表示可能なデータは公開用 API により取得できるデータで確認できる。地方自治体管理橋梁の点検調書は API で取得できないため、このページから個別ダウンロードする。

図-1 に概観を示す。橋梁の特定を防ぐため、マスク処理を施している。JSON ファイルは 1 橋梁につき大変数 shisetsu_id (橋梁 ID) から始まり、大変数 ikan_jyoukyou (移管状況) 内小変数 ikan_jiki (移管時期) で終わる。これが橋梁数分、繰り返される。本研究では特定地域の橋梁で、橋長 5.3m~5.5m の橋梁 60

橋分を JSON ファイルとしてダウンロードした。このうち、JSON ファイル (以後、API データ) と点検調書に齟齬があった橋梁 1 橋 (以後、検証橋) の点検調書のデータを検証用とする。

点検調書のデータ形式はエクセルデータであり、道路橋様式 1P001 というシートから始まる様式 1 を用いる。本シートでは橋梁名や橋梁 ID、橋長等の諸元、判定区分等が記載されている。エクセルデータのため、橋梁名等の変数名に対して変数値のセル位置が統一されていない点に留意する。例えば緯度・経度は変数名セルの右セルに変数値が記入されているが、橋長や橋幅などは変数名セルの下に変数値が記入されている。本研究ではこれらの関係を明示しなくとも ChatGPT で変数名と変数値の関係を整理可能かについても検証する。検証用の点検調書は検証橋のデータのみを用いる。Claude3 はエクセルデータを直接読み込むことはできないため、CSV 形式で保存しなおしたファイルを使用することとした。

API データと点検調書を Claude3-Sonnet で読み込ませ、表-1 にある対応関係を示し、それぞれ対応する変数を示すように指示する。変数の対応関係は既往研究を参考にした。この時、点検調書の変数名に対応する変数値のセル関係を明示せず指示する場合と、変数名に対応する変数値のセル番号を明示した場合とで実験を行う。前者が可能であればより簡単な指示で ChatGPT は点検調書のデータ構造を理解・分析できると考えられる。ただし、後者であっても点検調書のデータ様式が大きく変化した場合でもセル番号の対応関係を再度提示すればよいだけである点に留意する。

次に、得られた変数名と変数値の対応関係に対し、明確に異なる点を検知するよう指示する。これら不一致データについては既知である。この時得られた結果について定性的に分析し、本提案の可能性について検証する。

キーワード 道路橋デジタルツイン, 大規模言語モデル, Claude3, xROAD

連絡先 〒111-8648 東京都台東区浅草橋 5-20-8 CS タワー3F

技術創発研究所 高橋悠太 TEL 03-5822-2903

表-1 点検調書と API データの対応関係

No.	点検調書	API データ
1	橋梁名	shisetsu (meisyou)
2	橋梁 ID	shisetsu_id
3	緯度	ichi (ido)
4	経度	ichi (keido)
5	管理者名	kanrisya (meisyou)
6	路線名	rosen (meisyou)
7	定期点検 実施年月日	kiroku (jisshi_nengetsu)
8	緊急輸送道路	douro_kinkyu_syubetsu

No.	点検調書	API データ
9	代替路の有無	daitairo_umu
10	自専道 or 一般道	jisendo_mataha_ippando
11	占用物件 (名称)	senyoubukken_mei
12	判定区分	kiroku (hantei_kubun)
13	架設年次	kasetsu_nendo
14	橋長	kyouchou
15	幅員	fukuin
16	橋梁形式	kyouryou_keishiki



図-1 Claude3 : 点検調書と API データの対応関係



図-2 不一致データの検知結果

3. 結果と考察

出力された TXT ファイルと CSV ファイルで対応する変数関係を図-1 に示す。橋梁 ID について、対応関係を示せていないことがわかる。これは CSV ファイルにおける橋梁 ID が TXT ファイルと一致しないうえ、セルの結合等が原因で CSV 変換時に変数名セルから離れた位置に変数値が入ったこと原因と考えられる。

次に、橋梁 ID のセル番号を指定し、不一致データについて出力させた結果を図-2 に示す。既往の研究と比較し、この結果を出力するまでいくつか直接的なプロンプトを介した。おおむね不一致データを検知できていることが分かった。

4. まとめと今後の課題

ChatGPT を用いた既往の研究と比較し、エクセルデータを直接読み込めない、変数の不一致について

いくつかプロンプトを介した点などが欠点であった。ただし、回答については ChatGPT と比較して圧倒的に短時間で、セル番号をほとんど指定せずデータを読み込めるなど、優れた点も見受けられた。

今後の検討として、今回の CSV を ChatGPT で読み込み・検証し、ChatGPT と Claude を詳細に比較分析することが考えられる。

参考文献

- 1) 総務省：令和5年度情報通信白書、
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/nd247530.html>
- 2) 高橋悠太：ChatGPT を用いた xROAD 道路橋データの不一致箇所検知可能性の検証、AI・データサイエンス論文集、2024。(採択済み)
- 3) Claude3: <https://claude.ai/chats>
- 4) ChatGPT: <https://chat.openai.com/>
- 5) xROAD: <https://road-structures-db.mlit.go.jp/>