xROAD 道路橋データに対する ChatGPT の活用法検証

The verification of ChatGPT usage for xROAD road bridge data

高橋　悠太[*1]
Yuta Takahashi

[*1] 八千代エンジニヤリング株式会社
Yachiyo Engineering Co., Ltd.

Data in the civil engineering field is often language-based, and many drawings and other materials are only image data and have not been digitized. There is a lack of digital human resources to process these data, and the amount of data is enormous. This study verified the feasibility of using chatbots based on large-scale language models to process these data. ChatGPT can analyze the different format data in xROAD with some process, and the feasibility for data format robust processing tool is appeared.

1. 背景

土木構造物をデジタル空間にデータとして表現することは発災後の変化を捉えるなどに加え，日常的な点検・修繕を効率化することが期待されているが，目視点検を前提とする様々な規制により事業化が困難で，実現には至っていない．規制緩和に向けた取り組みとして，現段階で活用できるデータの付加価値を高めることで，課題や技術的成熟度を示すことも考えられる．これら取り組みはデータプラットフォームと活用法が求められる．

xROAD は国土交通省によってデジタル道路地図等を基盤として各種データを紐付けるデータプラットフォームとして構築されている[国土交通省 2022]．一環として，「民間企業等による技術開発の促進」，「これによる維持管理の更なる効率化」等を目指し，「全国道路施設点検データベース」の整備を進めている．2022 年 5 月に諸元・点検結果等の基礎データが無料で，同 7 月に詳細なデータを有料での公開を開始した．これらのデータは公開用 API を用いるなどして JSON 形式のファイルとして得られる[橋梁調査会]．ただし，2024 年 2 月現在では地方自治体の管理橋梁について，上記 API による点検調書等詳細なデータの取得はできていないが，点検調書については，HP 上で Excel データを直接ダウンロードできる．

データ形式は今後 JSON 形式に統合されていくことが予想されるが，その過程に合わせてアルゴリズムやシステムを構築することは Cost-Labor であり，不足が予測される人的デジタルリソースを必要とするため，データ形式の変化にロバストで簡易な分析手法が望ましい．本研究では大規模言語モデル(LLM: Large Language Model) を用いたチャットボットである ChatGPT [ChatGPT]に注目する．LLM を利用したチャットボットはマルチモーダル化が進んでおり，ChatGPT は OpenAI 社[OpenAI]によって開発された GPT (Generative Pre-trained Transformer)シリーズ[Brown 2020 など]を搭載している．Transformer[Vaswani 2017]を採用し，インターネット上にある大量のデータを学習することで GPT シリーズは構築されていると考えられており，ChatGPT は通常の教師学習に強化学習によるファインチューニングを加えることでさらに自然な言語生成を可能としている．2024 年 2 月現在，有料版では GPT-4 を使用可能で，ファイル読み取り機能が標準で搭載されている．加えて，チャット形式でのコード生成，読み取ったファイル上のデータ分析ができる．すなわち，No Code かつ厳密に条件を設定せずともデータ分析が可能ともいえる．これは初歩的なデジタル人材でもデータ処理や分析が可能であることを示している．例えば，Plugin を用いた点検調書作成について既往研究[青島 2023]で検討されている．

データ形式の変更に応じた分析が可能となるとき，ChatGPT はデータの構造を位置関係あるいは意味的関係において理解していることが求められる．例えば Excel データにおいて，変数名セルの近傍に変数値があるという位置関係がある．また，橋長と橋幅が近傍セルにあったとき，より適切な位置にある変数値を参照することが求められる．一方，意味的関係を理解している場合，橋長や橋幅には m などの単位がついた数字が変数値に入ることを理解していることが求められる．本研究では道路橋データの JSON ファイルと，対応する Excel データの点検調書を ChatGPT で分析し，変数の対応関係を理解しているか評価し，データ形式の変化にもロバストに分析可能か検証した．

2. データ形式と実験手順

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[xROAD]．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，個別にダウンロードした．

2.1 公開用 API で得られる JSON ファイル (API データ)

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1つの 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)の点検調書データを検証用とする．

2.2 点検調書

点検調書のデータ形式は Excel 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(図-2，本図は[国土交通省 2019]から引用し，著者が編集した)．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．

連絡先：高橋悠太，〒111-8648 東京都台東区浅草橋 5-20-8 CS タワー3F，yt-takahashi@yachiyo-eng.co.jp, 03-5822-2903

Excel 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．

2.3 変数の対応関係

本研究では基礎データ(橋梁位置や管理者，橋長等の物理諸元，判定区分等)を中心とした図-2 赤枠内の 16 変数について，フォーマット間で対応関係を設定し，**表-1**に示す．左列に点検調書の変数名，右列に JSON ファイルの変数名とする．右列括弧内は括弧外の大変数に連なる小変数名になる．この対応関係は ChatGPT を用いて点検調書で灰塗セルに変数名が格納されていることを利用して書き出し，API データの変数名との対応関係を出力させ，人による修正を加える形で設定した．

API データと点検調書のデータで形式的に異なる点として，いくつかの変数がカテゴリ変数によって設定されている．例えば，緊急輸送道路，代替路の有無，自専道 or 一般道については点検調書上では「有」などの文字データ，API データではカテゴリ番号を示す数値データになっている．

2.4 実験手順

API データと点検調書を ChatGPT 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．ChatGPT による JSON ファイルの分析は一般的に困難ではないが，点検調書は日本語を含み，一般的なデータ形式でないことから，点検調書における変数名と変数値が正しく抽出されているか注目する．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に指示する場合と，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形式が大きく変化した場合，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．API データ・点検調書の結果は一部異なるため，同じデータが抽出された場合確認可能である．結果を分析し，本提案の可能性について検証する．


```
{'shisetsu_id': '[redacted]',
 'shisetsu_kubun': '1',
 'koushin_nichiji': '2022-10-24T12:00:00+09:00',
 'nendo': '2023',
 'shisetsu_bangou': [redacted],
 'kanrisya_code': '[redacted]',
 'kiso_nendo': '2019',
 'sagyounaiyou_flg': {'kubun': '', 'uchiwake': '', 'uchiwakehosoku': ''},
 'syogen': {'kyouryou_keishiki': 'RC¥u3000T桁橋',
  'kasetsu_nendo': '不明',
  'kasetsu_suisoku_nendo': '',
  'kyouyou_nendo': '',
  'kyouryou_syubetsu': '',
```

.
.
.

```
  'ikan_joukyou': {'miikan_kyouryou': '',
   'miikan_kyouryou_kubun': '',
   'yousei_surubeki': '',
   'honrai_kanrisya_flag': '',
   'honrai_kanrisya_mei': '',
   'dourohou_douro': '',
   'ikango_youto_syubetsu': '',
   'genkyou': '',
   'ikan_jiki': ''}}},[EOF]
```


図-1　xROAD の公開用 API を用いてダウロード可能な JSON 形式のファイル概観

図-2　xROAD 上で取得可能な点検調書，道路橋様式 1

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	roka_jyoukyou (kinkyuu_yusou_douro_umu)
9	代替路の有無	daitairo_umu
10	自専道 or 一般道	rosen (douro_syubetsu)
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

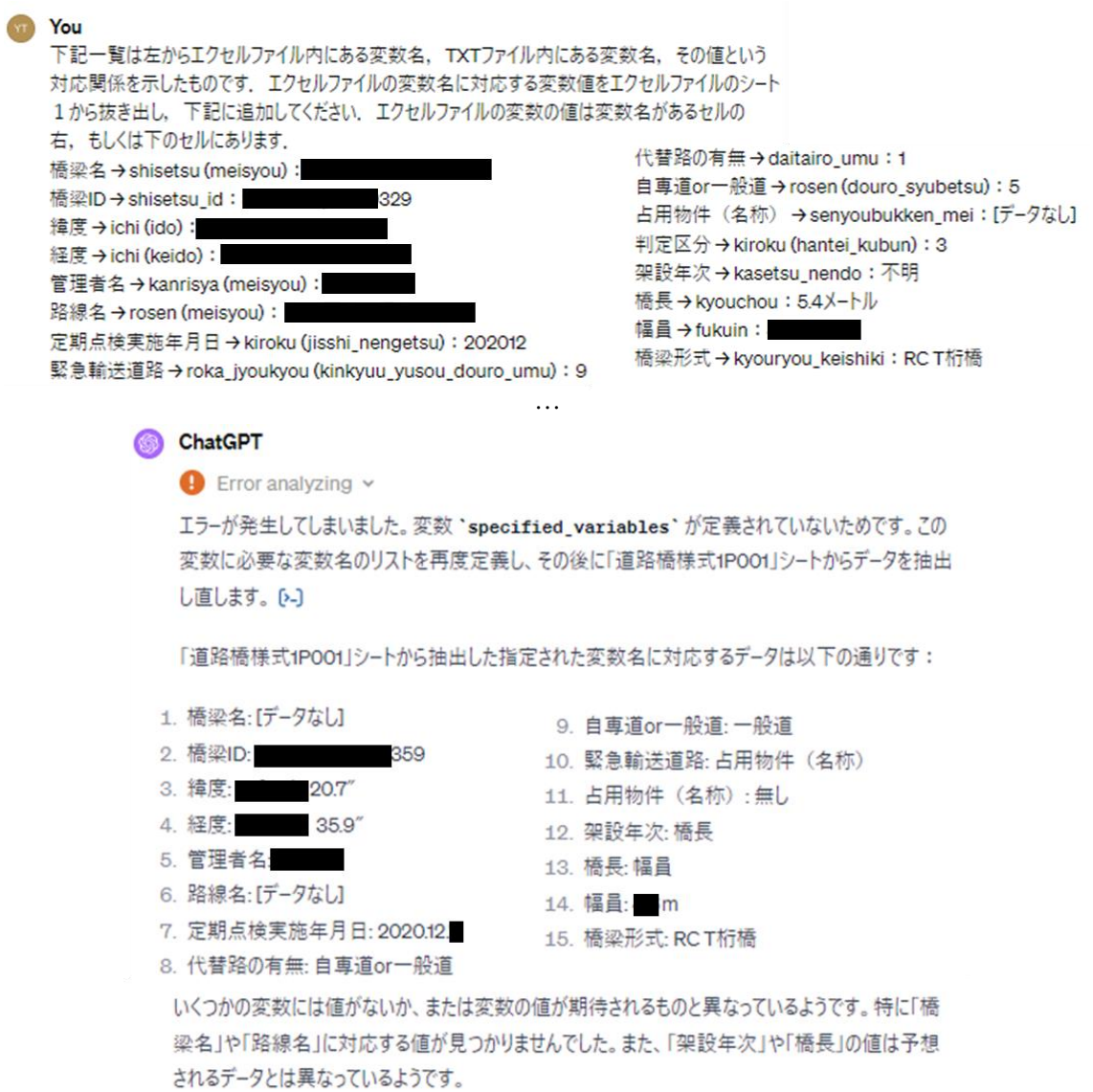

図-3　点検調書の変数値セルを明示しない場合

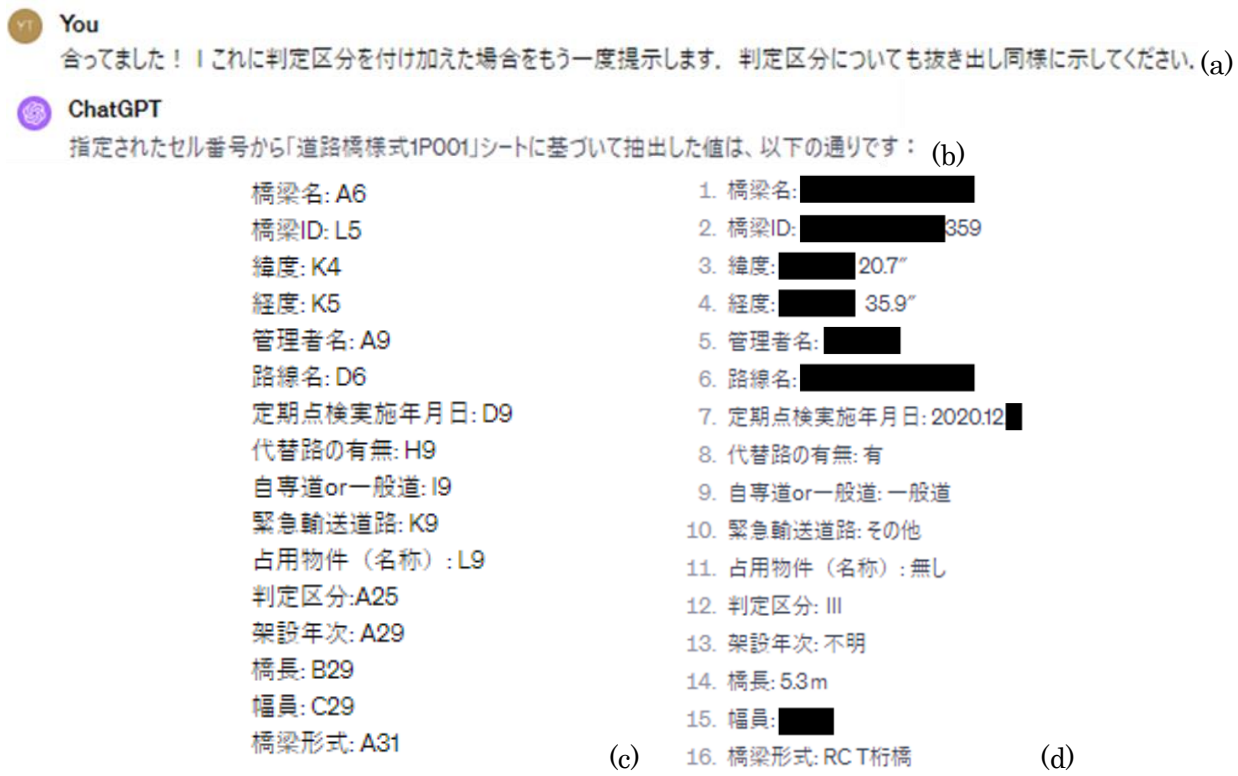

図-4　点検調書の変数値セルを明示した場合の対応関係

3. 実験結果と考察

ChatGPT 上で行った応答を画面キャプチャーした結果を図-3 および図-4 で示す．プロンプト自体が重要であることから，転記するのではなく図として示す．なお，すべての結果について，橋梁が特定できないよう，マスク処理を行うとともに，掲載に適するように画像編集を行っているが，結果には一切影響ない．

3.1 点検調書の変数値セルを明示しない場合

点検調書の変数値セル番号を明示せず，点検調書の変数名に対応する変数値を抜き出して示すように指示した結果を図-3 に示す．変数名と変数値との予想される関係のみでは正しいセルを参照できていないことがわかる．このことから現バージョンにおいては点検調書の変数名に対応する変数値セル番号を明示する必要があることが確認された．ただし，架設年次や橋長に適切なデータが参照されていないことを理解しており，今後の改良や異なる検討によって解決可能である可能性が示された．

3.2 点検調書の変数値セルを明示した場合

明示した場合について，点検調書の変数名と変数値の対応関係出力結果を図-4 に示す．図-4(a), (b)は人側からのプロンプトと ChatGPT の応答を示し，図-4(c), (d)はそれぞれ人側が与えた変数値セル番号，ChatGPT から出力された変数値になる．変数値セル番号は本実験で用いたファイルでの条件であり，フォーマットが更新される場合対応が必要となる．変数名に対し，適切な変数値を選択することができている．

4. まとめ

xROAD 道路橋データに対し，データ形式の変化にもロバストに処理・分析する方法として ChatGPT が適用可能か検証し，与えた JSON ファイルのデータに基づき点検調書(Excel データ)のデータ構造についても変数値セルを明示することで，JSON と Excel という異なるデータ形式であっても分析できる可能性が示された．今後は本検討を進め，ファイル間の対応関係を明示することなく，データ形式間で対応する変数関係を抽出可能な手法について提案・検証する．

参考文献

[国土交通省　2022]　国土交通省: 道路施設の詳細な点検データの公開開始, 2022.

[橋梁調査会]　橋梁調査会: 全国道路施設点検データベース（道路橋）公開用 API 仕様書, 77 条調査データ編. https://www.jbec.or.jp/upfile/road/data03/%E5%85%AC%E9%96%8B%E7%94%A8API%E4%BB%95%E6%A7%98%E6%9B%B8(%E9%81%93%E8%B7%AF%E6%A9%8BDB_77%E6%9D%A1%E8%AA%BF%E6%9F%BB%E7%B7%A8).pdf

[ChatGPT] ChatGPT: https://chat.openai.com/

[OpenAI] OpenAI: https://openai.com/

[Brown 2020 など] 例えば，Brown, T. B. et al: Language Models are Few-Shot Learners, arxiv, 2020. https://arxiv.org/abs/2005.14165

[Vaswani 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I.: “Attention is all you need”, NIPS 2017, 2017. https://arxiv.org/abs/1706.03762

[青島　2023]　青島 亘佐, 宮内 芳維: 大規模言語モデルの活用による橋梁点検調書作成の省力化に関する検討，AI・データサイエンス論文集，4 巻 3 号，p. 274-284，2023.

[xROAD] xROAD: https://road-structures-db.mlit.go.jp/

[国土交通省 2019]　国土交通省: 記録様式作成にあたっての参考資料（道路橋定期点検版），2019. https://www.mlit.go.jp/road/sisaku/yobohozen/tenken/yobo4_1-4.pdf