LLM ベースと従来型ジオコーディングの精度比較：日本の大・中・小規模地域における評価

卢　诗琪*・龐　岩博・関本　義秀**

Comparing the Accuracy of LLM-based and Geocoding with Traditional Methods: Evaluation across Large, Medium, Small Regions in Japan

Shiqi Lu*, Yanbo Pang**, Yoshihide SEKIMOTO

Abstract: In recent years, large language models (LLMs) have shown promise for geocoding, i.e., deriving geographic coordinates from textual place descriptions. While traditional methods, such as the Google Maps Geocoding API, achieve high accuracy for structured addresses, LLMs can better interpret unstructured data, though their geocoding accuracy remains uncertain. In this study, we systematically compare LLM-based geocoding with traditional methods using Zenrin Telepoint data as reference coordinates. Across representative large, medium, and small regions in Japan, LLMs show competitive performance, achieving mean errors of 0.94 km, 2.03 km, and 4.85 km, respectively, with an overall average processing time of 1.65 seconds. These results demonstrate that LLMs can offer comparable or improved geocoding accuracy relative to baseline methods, while maintaining efficiency.

Keywords: 大規模言語モデル（Large Language Models, LLMs），ジオコーディング（Geocoding），位置情報解析（Geographic information analysis），精度比較（Accuracy comparison），日本の地域（Japanese regions）

１．はじめに

近年，LLM（大規模言語モデル）の進展により，地理空間情報の自動生成・理解を前提とする応用（デジタルツイン，都市 OS，政策立案支援など）が急速に広がっている．しかし，その信頼性評価，とりわけ「位置が正しいか」を問う基礎能力の検証は十分とは言えない．高精度な空間参照が担保されなければ，意思決定への適用はリスクとなる．本研究は，LLM が地理空間情報をどこまで理解し，実座標を正しく復元できるのかを検証する最初の一歩として，ジオコーディングに焦点を当て，従来型手法と比較しながら精度・頑健性・処理時間を体系的に評価する．

１.１　研究の目的

本研究の目的は, LLM を用いたジオコーディング手法と従来型手法を，日本の大・中・小規模地域における多様な住所表現に適用し，その位置推定精度，頑健性，および処理時間を体系的に評価・比較することである．本研究により，各手法の精度特性と限界を明らかにするとともに, LLM の地理空間情報理解能力を実座標復元の観点から検証し，従来型手法との補完的関係や実務応用上の可能性を示すことを目的とする．

１.２　先行研究

地名解決（トポニム解決）とジオコーディングに関する先行研究では，主に従来型手法と大規模言語モデル（LLM）を活用した手法の発展が確認される．

従来の地名解決手法では，ルールベースや統計的手法が広く用いられてきた．例えば，ルールベースのアプローチでは住所体系の階層構造を利用して文字列マッチングを行う方法があり (Zandbergen, 2009)，統計的手法では隠れマルコフモデル（HMM）や条件付き確率場（CRF）が地名のシーケンスタギングに応用されている．近年では, 双方向 LSTM-CRF モデルが高い性能を発揮し, 住所要素の抽出精度を向上させたことが報告されている）(Zhengcong Yin, 2023).

一方, Transformer ベースのモデル (BERT, roBERTa など）や LLM（GPT-3 など）の応用が注目されている．GENRE や BLINK といったエンティティリンキングモデルは, Wikipedia などの大規模知識ベースを活用して地名の曖昧性解消を試みているが，非構造化テキストや誤入力に対する頑健性に課題が残る．また，GPT-3 を含む LLM は自然言語理解能力に優れるものの, 内部に地理データベースを持たないため，高精度な座標出力が難しいことが明らかにされている (Xuke Hu, 2024).

近年の研究では, LLM と従来型手法の融合が提案されている．Hu（2024）は，軽量なオープンソース LLM（Mistral，Llama2 など）を微調整し，地名の明確な参照先（市区町村，国など）を推定した後，GeoNames や ArcGIS などのジオコーダーで座標を取得するハイブリッド手法を開発．これにより，従来の投票システムや GENRE よりも高精度（Accuracy@161km で 0.91）かつ計算効率に優れる結果を示した．

*　学生会員　東京大学大学院新領域創成科学研究科（The University of Tokyo）
〒153-0041　東京都目黒区駒場 4-5-29　E-mail：6056775605@edu.k.u-tokyo.ac.jp
**　正会員　東京大学空間情報科学研究センター（The University of Tokyo）

２．検証対象と研究方法

本研究では，まず川崎市をテスト地域として選定し，100 件の実住所サンプルを抽出して，現行の GPT モデルのジオコーディング能力を検証した．ジオコーディング過程における課題を解決し精度を向上させるため，適切なプロンプトを設計し，処理時間も記録した．その後，最も性能の良いプロンプトを用いて，日本国内の大規模・中規模・小規模の各 5 地域（合計 15 地域）における GPT モデルのジオコーディング能力を評価した．

２.１　使用データとベースライン設定

本研究では，日本国内の大・中・小の規模別に分類された各 5 地域，合計 15 地域を対象とする（表 1 参照）．各地域からランダムに 100 件の実住所サンプルを抽出し，区・町丁目・番地レベルまでの詳細情報を含むデータを収集した．

表 1　対象地域表

地域規模	都道府県・市町村名
大規模	東京都，大阪府，横浜市，札幌市（北海道），福岡市（福岡県）
中規模	金沢市（石川県），松山市（愛媛県），熊本市（熊本県），高松市（香川県），盛岡市（岩手県）
小規模	白川村（岐阜県），与那国町（沖縄県），大間町（青森県），小豆島町（香川県），鶴居村（北海道）

本研究における正解座標は，ゼンリンテレポイントデータを用いた．これは全国の住所を号・番地レベルまで高精度に位置付け可能な座標データである．

さらに，従来型ジオコーディング手法との比較のため，Google Maps API による住所変換結果をベースラインとして設定し，LLM 出力との比較を行った．

２.２　使用モデル

本研究では，Azure OpenAI の GPT-4，GPT-4o，および GPT-O1 / GPT-O1 Mini を対象に，地理座標推定能力を検証した．初期検証の結果，GPT-4o は精度面で最も優れていたが，異なる住所に対して同一座標を出力する重複問題が確認された．一方，GPT-O1 および GPT-O1 Mini では，この重複問題は観察されなかった．

そこで，GPT-4o に対してプロンプト設計の改善を行い，重複問題の解消と精度維持を確認した．最終的に，本研究の正式テスト段階では，改善済みプロンプトを適用した GPT-4o を使用した．

２.３　プロンプト設計

本研究では，LLM に対するジオコーディングの指示文（プロンプト）を設計し，精度向上および出力の安定化を図った．初期のプロンプトでは GPT-4o において 59％の重複座標が出力されたが，最終的なプロンプトでは 0 対に低減され，重複問題の解決が確認された．最終的に改善したプロンプトの特徴は以下の通りである．

1)　住所の階層的解析指示

都道府県，市区町村，町丁目，番地・号といった住所の階層を逐次解析するよう明示し，各階層を対応させることで住所の理解を促進した．

2)　精度の明示的制約

出力される緯度・経度の精度を少なくとも小数点以下 6 桁以上と指定し，高精度な座標生成を保証した．

3)　出力形式の厳格化

座標および付随情報の出力形式を厳密に指定し，解析・可視化処理への適用を容易にした．

4)　反復性制約の導入

過去に出力された座標を再利用しないよう制約を設けることで，不確実性や重複座標の問題を低減した．

5)　Few-shot 学習の活用

簡単，中程度，複雑な住所のサンプルを提示する few-shot 形式により，モデルが多様な住所表記に対応できるよう誘導した．

これらの改良により，最終プロンプトを用いた GPT-4o において，精度向上と重複座標の低減が確認された．

基本プロンプト：

> I have a place in Japan: "{place_name}".
> Please provide the approximate latitude and longitude in decimal degrees.
> Format your answer as: "Latitude: xx.xxxx, Longitude: yy.yyyy".
> If you don't know, make your best guess.

改善済みプロンプト：

> You are tasked with geocoding a Japanese address to latitude and longitude coordinates.
> Instructions:
> Read and analyze the entire Japanese address carefully, including all hierarchical levels.
> In Japan, addresses generally follow this order:
> 「都道府県 → 市（市区町村）→ 区 → 町 → 丁目 → 番地 → 号 → 建物名・部屋番号」.
> Note:
> - For 政令指定都市 (designated major cities such as Yokohama, Kawasaki, Osaka), "区 (ku)" refers to administrative wards under the city.
> - For 東京都特別区部 (Tokyo 23 Special Wards, e.g., 新宿区, 港区), each "区" functions as a municipality itself, equivalent in status to a city.
> Use all details provided (from prefecture down to the smallest available unit, such as block, building number, or apartment room).
> Provide coordinates that best match the complete address provided.

Do not truncate or ignore any part of the given address; use all provided details to achieve maximum accuracy.

Here, maximum accuracy means:

Latitude and longitude must be reported with at least 6 decimal places (≈ within 0.1 meter precision).

All hierarchical components of the Japanese address must be analyzed.

The output coordinates should correspond to the smallest available unit of the address, not a higher-level centroid (e.g., not just the ward or city center).

Avoid duplicating coordinates from previous queries; treat each address as unique.

If uncertain, provide your best possible guess, clearly based on the complete address.

Output Format:

Exactly format your response as follows:

Latitude: xx.xxxxxx, Longitude: yy.yyyyyy

Do NOT use any Markdown formatting such as ** or *; output plain text only.

Examples:

Example 1:

Address: 東京都　目黒区　青葉台　１丁目　１６－９

Latitude: 35.647322, Longitude: 139.696684

Example 2:

Address: 静岡県　静岡市葵区　相生町　１４－２８－１Ｆ

Latitude: 34.977019, Longitude: 138.392405

Example 3:

Address: 岐阜県　大野郡白川村　大字荻町　２４９９

Latitude: 36.255081, Longitude: 136.9023

Example 4:

Address: 京都府　京都市中京区　木屋町四条上

Latitude: 35.003803, Longitude: 135.770556

Now, please geocode the following address with careful attention to the full details provided:

Address: "{place_name}"

Provide the coordinates below:

2.4　評価指標

本研究では，従来型手法（Google Maps API）をベースラインとし，LLM 出力との相対比較によって性能を評価する．評価指標は以下の 3 点である．

1)　座標精度（距離誤差）

モデルが出力した緯度・経度およびベースライン（Google Map API による座標出力）を，参照座標（ゼンリンテレポイントデータ）と比較し，各モデルおよび地域ごとの誤差分布，中央値，標準偏差を算出・評価した．

2)　処理時間

各モデルが住所入力から座標出力までに要する平均処理時間（秒）を計測し，モデルの計算効率を評価した．

3)　座標の一意性（重複座標の有無）

出力された座標が異なる住所に対して同一座標となる事象の発生を確認し，モデルの出力の一意性・頑健性を評価した．

3. 実験結果

3.1　予備実験

本研究では，最終的に GPT-4o を採用する前に，複数のモデルを対象として精度，処理時間，およびデータ重複率の 3 つの指標で性能比較を行った．各モデルの結果は以下の通りである（表 2～表 4, 単位は km および秒）．

表 2　各モデルの数値指標

モデル	平均誤差	標準偏差	最小誤差
GPT-4	1.76	0.742	0.002
GPT-4o	1.041 ～ 1.289	0.683 ～ 0.830	0.000 ～ 0.001
GPT-O1	1.091	0.721	0.003
GPT-O1 Mini	4.417	3.595	0.002

表 3　各モデルの平均処理時間

モデル	平均時間（秒）
GPT-4	2.48
GPT-4o	1.83
GPT-O1	15.51
GPT-O1 Mini	4.73

表 4　各モデルの座標の一意性

モデル	繰返数
GPT-4	78%
GPT-4o	59%→0
GPT-O1	0
GPT-O1 Mini	10%

表 2～表 4 より，GPT-4o は平均誤差が最小で精度に優れると同時に，処理時間も短く，プロンプト改善によりデータ重複率を 0 に抑えることが可能であった．一方，GPT-O1 は精度は比較的高いものの，処理時間が非常に長く, GPT-O1 Mini は誤差が大きく不安定であった. GPT-4 は精度は中程度であるものの，重複率が高い点が課題であった．

以上より，本研究の本実験では，改善済みプロンプトを適用した GPT-4o を採用した．

3.2　本実験

3.2.1　平均精度の比較

図 1 に，日本大・中・小代表地域における Google map API と GPT によるジオコーディング精度の平均比較を示す．横軸は各地域，縦軸は平均誤差距離（km）である．

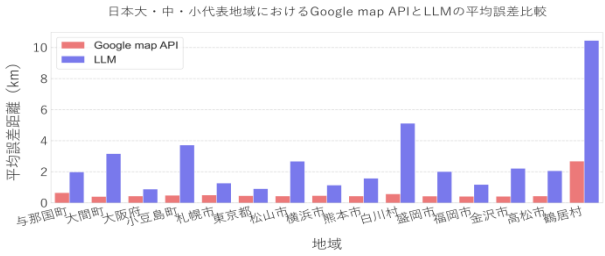


図 1　代表地域における平均誤差

図 1 より，Google map API は全地域で平均誤差が 1km 程度かそれ以下で，精度が安定して高い．GPT は東京都，大阪府など大都市では比較的小さい誤差を示すが，白川村や鶴居村など郊外・農村地域では平均誤差が 5km を超え，鶴居村では 10km を超えるなど，地域の規模や都市化の程度に応じて精度に顕著な差異が見られる．

3.2.2　誤差分布の詳細分析

図 2 は，同地域における Google map API と GPT の誤差分布比較を示す．横軸は各地域，縦軸は誤差距離（km）である．

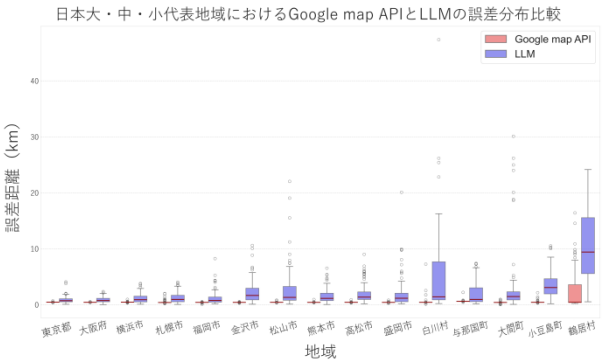


図 2　代表地域における誤差分布比較（箱線図）

図 2 より，Google map API は全地域で誤差が小さく，分布も比較的狭い．GPT については，東京都，大阪府，横浜市など大都市では誤差が小さく，分布も狭いが，白川村や鶴居村などの一部郊外・農村地域では誤差の分布が非常に広がり，大きな誤差も見られる．これにより，GPT の誤差分布は地域の規模や都市化の程度に応じて顕著な差異が存在することが示された．

3.2.3　処理時間および座標の一意性の比較

各地域における GPT-4o のジオコーディング処理時間およびデータ重複率を測定した．全地域の平均処理時間は 1.65 秒 で，概ね 2 秒以内に収まった．出力の一意性は多くの地域で確保されているが，与那国町では重複が顕著（18 件）であった．

4．考察

本研究では，GPT-4o をはじめとする LLM が日本国内の多様な地域において住所から座標を推定できることを示した．大都市圏では平均誤差が 1 km 前後と比較的高精度であった一方，白川村や鶴居村など一部の小規模自治体では誤差が大きく，また与那国町においては繰返し出力が多く観察されるなど，LLM の地理的常識・地名解釈能力には限界も確認された．これらの結果は，LLM の適用可能性が地域特性や訓練データの偏りに左右されることを示唆している．他方，適切なプロンプト設計を行うことで重複出力を回避でき，出力の安定性を改善できることが確認された点は重要である．また，従来のジオコーディング API が入力形式に依存しやすいのに対し，LLM は文脈理解に基づく柔軟な推定が可能であるため，両者を補完的に活用することで精度と頑健性の向上が期待できる．さらに，本研究で測定した処理時間はいずれの地域でも 2 秒以内に収まり，実用性の観点からも有望であることが示された．

5．結論と今後の課題

本研究は，LLM によるジオコーディングの可能性と限界を実証的に明らかにし，特に GPT-4o が従来手法を補完する形で活用し得ることを示した．加えて，プロンプト設計の工夫により精度と一意性が改善されることを確認し，LLM の柔軟性が地理情報処理において有効であることを示した点に意義がある．

今後の課題としては，第一に LLM のジオコーディング精度をさらに高める手法の探究，第二に GPT-4o 以外の最新 LLM を含むモデル間比較による最適モデルの検証，第三に LLM と既存ジオコーディング API を統合したハイブリッド手法の開発が挙げられる．また，少数ショット学習や追加情報（郵便番号・都道府県名など）の付与による精度向上の可能性についても検討が必要である．これらの課題に取り組むことで，LLM に基づく地理情報処理は実用的かつ信頼性の高い次世代技術へと発展することが期待される．

謝辞

研究室の皆様にご支援をいただき，感謝申し上げます．

参考文献

Zandbergen, P. A. (2009). Geocoding quality and implications for spatial analysis. *Geography Compass*, **3**(2), 647-680. DOI: 10.1111/j.1749-8198.2008.00205.x.

Yin, Z. C., Li, D., Goldberg, D. W. (2023). Is ChatGPT a game changer for geocoding - a benchmark for geocoding address parsing techniques. *In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Searching and Mining Large Collections of Geospatial Data*, 1-8. DOI: 2310.14360.

Hu, X., Kersten, J., Klan, F., Farzana, S. M. (2024). Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge. *International Journal of Geographical Information Science*, https://doi.org/10.1080/13658816.2024.2405182.