

LLM の土木・建設分野への利活用

菅田大輔・箱石健太・一言正之

本稿では、土木・建設分野における大規模言語モデル（LLM）の利活用を目的として、文書特化型チャットボットやテキスト埋め込み技術の利活用の観点から検証を行った。その結果、文書特化型チャットボットは、土木・建設分野の文書に即した正確な回答が可能であることを確認できた。また、テキスト埋め込み表現モデルは土木・建設分野のテキストに対しても適用性があることを確認した。さらに、土木・建設分野のテキストのベクトル値を学習した機械学習モデルを活用することで、高い精度で分類ができる可能性が示唆された。

キーワード：生成 AI、大規模言語モデル、テキスト埋め込み表現、機械学習、プロンプトエンジニアリング

1. はじめに

土木・建設分野には河川砂防技術基準といった技術基準書やマニュアル、ガイドラインなど、膨大な文書データが存在する。これらの文書データから必要な情報を迅速かつ正確に取得できることは、生産性向上の観点から重要である。

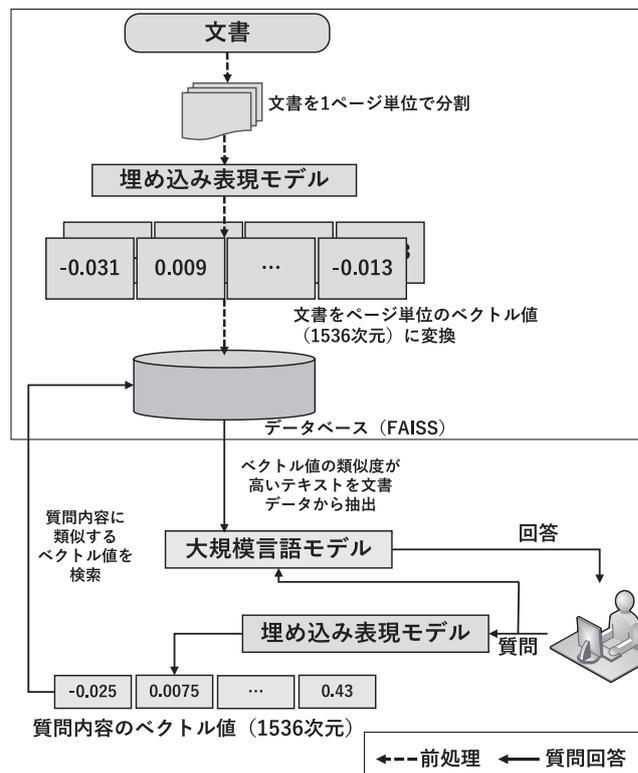
そのような中、ChatGPT を始めとする大規模言語モデル（以下、LLM）の質疑応答性能、埋め込み表現性能の高さは近年注目を集めている。土木・建設分野で扱われる専門用語の意味などを踏まえた LLM との対話により、効率的な情報抽出が可能となれば、先述した課題を解決できる可能性がある。

そこで、本稿では土木・建設分野で取り扱われる文書データに特化したチャットボットを構築し、質疑応答タスクの性能を検証した。また、質疑応答以外のタスクにおける LLM の適用性や精度も検証するため、公共工事に活用される複数の新技術¹⁾の概要文に対し、テキスト埋め込み技術を適用し、類似する情報抽出タスクの性能を検証した。さらに、土木・建設分野特有の文章が属する区分を分類するタスクを設定し、プロンプトエンジニアリングや機械学習モデルを用いて検証を行った。なお、全ての検証タスクに共通して用いた LLM やテキスト埋め込み技術は、Azure Open AI Service²⁾の提供するモデルとした。

2. 文書特化型チャットボットの利活用

(1) 文書特化型チャットボットの構成

LLM の知識は訓練データに限定されているため、LLM に問いかける質問の内容によっては事実に基づかない回答を生成する（ハルシネーションを発生する）可能性がある。本検証ではこのようなハルシネーション



図一 文書特化型チャットボットの構成

ンの発生を抑制するため、RAG (Retrieval-Augmented Generation：検索拡張生成)³⁾に基づく文書特化型チャットボットを構築した。具体的には、土木・建設分野の文書データを格納した外部データベースから質問内容に関連する情報を取得し、その情報を基に回答用の文章を生成する構成でチャットボットを構築した(図—1)。チャットボットの処理の詳細は次項に示す。なお、LLMはgpt-3.5 turboを使用した。

(a) 前処理

はじめに、文書データ(pdfファイル)を1ページ単位に分割する。次いで、テキスト埋め込み表現モデル(本稿では、Azure OpenAI Serviceが公開する埋め込み表現モデルの中で高性能なもの1つであるtext-embedding-ada-002を使用)を通じ、ページ毎に記載されたテキストを数値ベクトル(1536次元)に変換する。その後、数値ベクトルのインデックスをFAISSの提供するデータベースに格納する。なお、FAISSとはMeta社が開発したベクトルの類似検索・クラスタリングのためのライブラリである。

(b) 質問応答処理

ユーザが行う質問内容のベクトル値に対し、FAISSに格納されたベクトル値のインデックスから類似度が高いテキストを検索する。その後、LLMは、質問内容と抽出したテキストを結合したうえで、回答内容を生成する。なお、質問内容のベクトル値生成にはマニュアル同様、text-embedding-ada-002を用いた。加えて、LLMに事前に与えたインストラクションには「答えが分からない場合は、答えを作ろうとせず、分からない旨を回答すること」という内容を含めた。

(2) 検証結果と考察

前節の構成を基に、文書特化型チャットボットを構築した。また使用した文書は、NETISの独自ルールが記載された「公共工事等における新技術活用システム」申請マニュアル(令和4年3月)とした。

本検証では、当該マニュアルの記載内容を踏まえた質問内容を作成し、回答を生成した。表—1に示す生成結果より、総じて申請マニュアルの記載内容に即した正確な回答ができていることが確認できた。また、敢えて申請マニュアルに記載されていない質問を行ったところ、「詳細については文書中には触れられていないため、正確な回答をすることはできません。」という回答が生成された(表—1の番号3)。このように回答生成に必要な情報を参照できないケースでは、チャットボットが回答を生成しないことを確認した。これは、チャットボットで事前に設定したインストラクションに「答えが分からない場合は、答えを作ろうとせず、分からない旨を回答すること」と記述した内容が反映されているものと考えられる。

(3) 今後の課題

本稿では、生成内容の正確性を定性的に評価したが、今後は生成内容の妥当性を評価するための定量的かつ適切な指標が必要となる。また、本稿では質問内容に関連する情報を検索するために、ベクトル検索を用いた。生成内容の正確性を更に高める手法として、HyDE(Hypothetical Document Embeddings)に基づくベクトル検索、全文検索とベクトル検索を組み合わせたハイブリッド検索、セマンティックハイブリッド検索など複数の検索手法が提案されている。検索手法を変更した場合の精度検証も今後課題と考えられる。

表—1 NETIS マニュアルに関する質問及びその回答内容

番号	質問内容	回答内容
1	NETISに登録されている評価情報とはどのような情報ですか。	NETISに登録されている評価情報は、新技術を活用して行われた公共工事等での事後評価結果に関する情報です。具体的には、新技術の優位性、安定性、現場適用性、技術特性などを総合的に評価した結果が含まれます。また、評価情報には、事前審査の結果に関する情報も含まれます。
2	NETIS(評価情報)の掲載期間を教えてください。	NETIS(評価情報)の掲載期間は、掲載された日の翌年度の4月1日から起算して5年を経過した日までとされています。ただし、活用効果評価が実施されてNETIS(評価情報)に反映された場合は、掲載期限が変更され、NETIS(評価情報)に反映した日の翌年度の4月1日から起算して5年を経過した日までとなります。
3	NETIS登録番号には何か意味があるのですか。	文脈から判断する限り、NETIS登録番号は技術に対して付与される番号であり、ある種の識別子として機能しているようです。詳細については文書中には触れられていないため、正確な回答をすることはできません。

3. テキスト埋め込み技術の活用

(1) テキスト埋め込み技術とは

テキスト埋め込み技術とは、テキストの意味を多次元の数値ベクトルに変換する技術である。一般的に類似した意味を有するテキストがある場合、それらは近接したベクトルとして表現される。

本稿では、LLM が土木・建設分野のテキストから類似する情報を抽出できるかを検証するため、NETIS で公開されている複数の新技術の説明文に対してテキスト埋め込み技術を適用した。NETIS 利用を想定した場合、利用者が閲覧中の新技術に類似する新技術を、テキスト埋め込みを通じて推薦表示できれば、公共工事に活用する新技術の比較検討の支援が期待できる。

(2) 検証の流れ

テキスト埋め込みを用いた新技術のベクトル化の検証フローを 図-2 に示す。はじめに、新技術情報を説明したテキストを 1 件ずつ、テキスト埋め込み表現モデル（本稿では、text-embedding-ada-002 を使用）を介し、ベクトル値に変換する。ベクトル値は 1536 次元の多次元情報であり、このままでは意味の類似する技術情報が近接配置されているかを検証できない。そこで、次元圧縮アルゴリズムの t-SNE (t-Distributed Stochastic Neighbor Embedding)⁴⁾ を適用し、2 次元にデータを変換の上、その分布を可視化する。なお、新技術を説明するテキストは、技術名称、アブストラクト、概要、新規性及び期待される効果、比較する従来

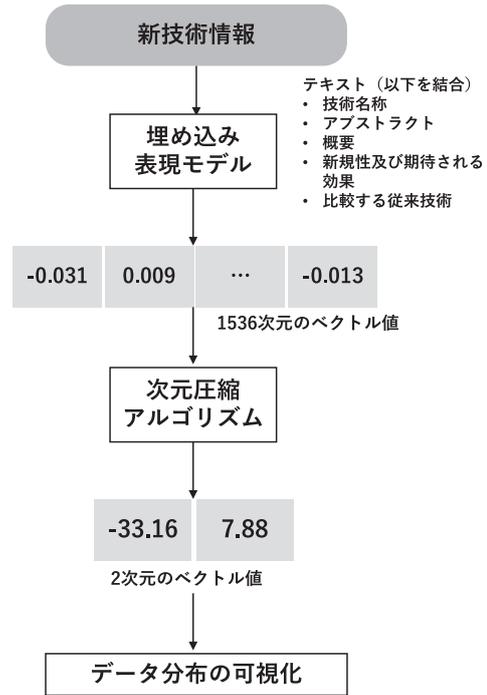


図-2 テキスト埋め込みを用いた新技術のベクトル化の検証フロー

技術を結合したものとした。

(3) 結果と考察

新技術情報の全件を図化した結果を図-3 に、図-3 の枠で囲った抽出事例を図-4 と図-5 に示す。図-4 では LED 照明灯や警告灯といった照明関連の新技術が近接して配置されていることが分かる。また、図-5 では遠隔監視や遠隔臨場、連絡ツール等、遠隔操作に関する新技術が近接して配置されているこ

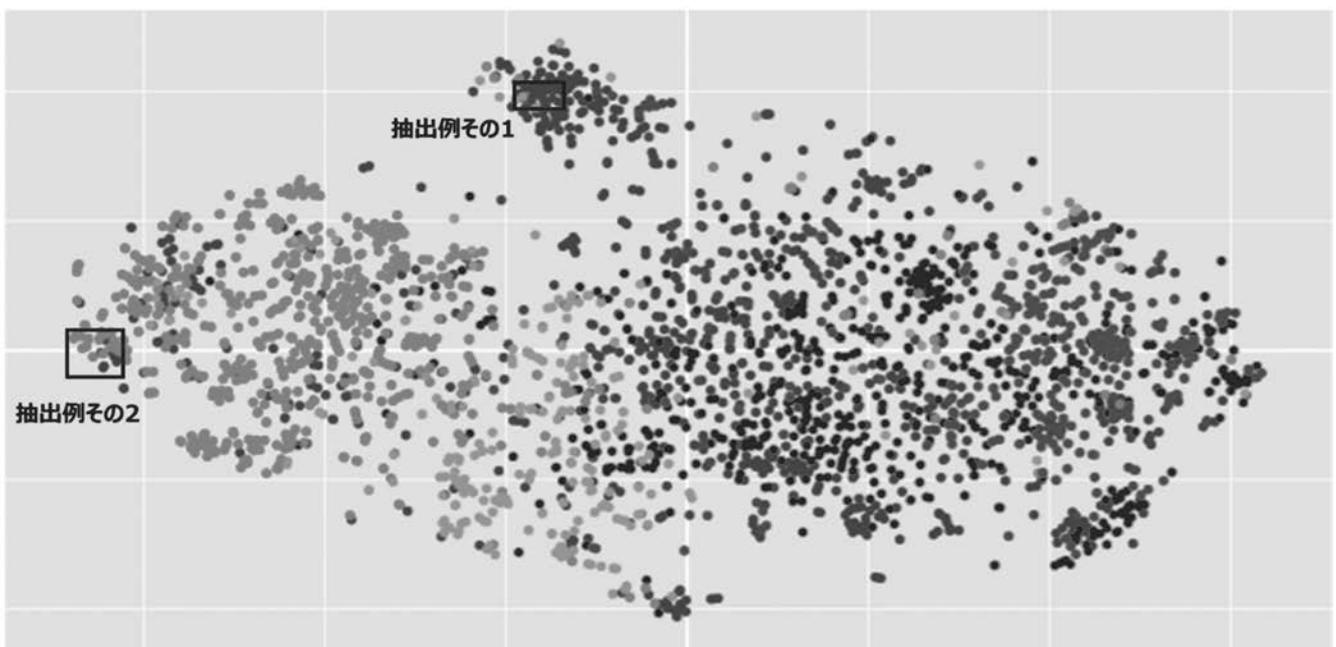


図-3 新技術情報の散布図 (全体像)

した。本検証では NETIS 登録された新技術情報が属する区分を分類させることとし、機械学習モデルには GBDT (Gradient Boosting Decision Tree : 勾配ブースティング決定木) で高い精度が報告されている Catboost⁵⁾ を採用した。

(1) 検証の流れ

機械学習モデルによる分類の流れを述べる。NETIS 登録された新技術の抽象化に対しテキスト埋込みを適用し 1536 次元に変換した数値ベクトルを機械学習モデルの訓練データとし、新技術の区分(製品, 工法, システム, 材料, 機械) を分類した。分類に際しては 5 分割交差検証を実施した (データ件数のうち 5 分の 4 が訓練データ, 残りがテストデータの構成)。本稿で検証したデータは, 区分ごとにデータ件数の偏りがあったため, 層化抽出法により分割データ内の区分割合を揃えた。

また, プロンプトエンジニアリングにより LLM で分類させるケースでは, プロンプトと新技術の抽象化を LLM の入力とし, 複数のプロンプトを設定した。具体的には, プロンプト内に回答事例を記述しない zero-shot learning, 回答事例を数例記述する few-shot learning, ペルソナ (職務内容や役職) などの設定を組合せた。

(2) 結果と考察

検証結果を表一 2 に示す。表一 2 の番号 1 ~ 3 で

表一 2 プロンプトエンジニアリング及び機械学習による土木・建設分野の文章の分類結果

番号	条件	正解率
1	zero-shot learning, ペルソナ設定無し	61.8%
2	zero-shot learning, ペルソナ設定有	63.9%
3	few-shot learning, ペルソナ設定有	68.2%
4	機械学習	73.0%

用いたプロンプトは表一 3 に示す。

プロンプトエンジニアリングによる分類を実施した番号 1 ~ 3 に着目すると, ペルソナを設定し, 且つ few-shot learning を適用した事例が最も分類精度の高い結果となった。また, 番号 4 の機械学習に着目すると, プロンプトエンジニアリングよりも, GBDT を用いた分類結果が最良であった。プロンプトエンジニアリングは, 本稿で取り上げた手法以外にも Chain-of-Thought プロンプティング等, 様々な手法がある。今回の検証では, プロンプトエンジニアリングよりも GBDT を用いた分類のほうが総じて高い結果となったが, プロンプトの設定次第では機械学習による分類を上回る可能性も考えられる。この点に関する検証は今後の課題としたい。

5. おわりに

本研究では, 土木・建設分野で取り扱われるテキストを用意し, LLM の利活用に向けた精度検証を行った。その結果, 土木・建設分野においても, 文書特化

表一 3 プロンプトの詳細

番号	条件	プロンプト
1	zero-shot learning, ペルソナ無し	次の文章を 1. 工法, 2. 材料, 3. 機械, 4. 製品, 5. システムのいずれか 1 つに分類してください。なお, 回答は番号のみで答えてください。また, 分類の説明は以下の通りです。 ・「工法」: 材料, 機械, 製品, システム等を組み合わせることにより, 工事の一部, あるいは全体を完遂させるに足る方法。 ・「材料」: 公共工事等において, 加工, 添加等されることにより用を成す原料, 資材。 ・「機械」: 公共工事等に用いる建設機械, 作業用機械。 ・「製品」: 公共工事等により製作, 築造される物を構成する一部材であり, 新たな加工を要さないもの。 ・「システム」: 材料, 製品, 機械等が体系的に組み合わせられ, 公共工事等における工法, 調査方法等を支援するものや, IT 等の先端技術を利用した情報システム, 施工管理や維持管理に利用するマネジメント技術などを指す。
2	zero-shot learning, ペルソナ有	あなたは, 優秀な土木技術者で, 土木建設分野の新技術に関する文章から, その新技術がどのジャンルに分類されるか推定することができます。 {以下, 番号 1 のプロンプト同様。}
3	few-shot learning, ペルソナ有	{番号 2 のプロンプト同様} {工法に関する抽象化} {工法} {材料に関する抽象化} {材料} ・ ・ {システムに関する抽象化} {システム}

型チャットボットやテキスト埋め込み表現の有効性が確認された。また、テキスト埋め込みによるベクトル値を入力とした機械学習モデルは、本検証におけるプロンプトエンジニアリングによる分類よりも精度が高いことを確認した。



《参考文献》

- 1) 国土交通省：新技術情報提供システム（NETIS）ウェブページ
(<https://www.netis.mlit.go.jp/>)
- 2) Microsoft: Azure OpenAI Service ウェブページ
(<https://azure.microsoft.com/ja-jp/products/ai-services/openai-service/>)
- 3) Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela, Advances in Neural Information Processing Systems 33, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, 2020.
- 4) Maaten, L. and Hinton, G.: Visualizing data using t-SNE, Journal of machine learning research, pp.2579-2605, 2008.
- 5) Prokhorenkova, Liudmila, et al. CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems. 2018.

【筆者紹介】

菅田 大輔 (すげた だいすけ)
日本工営(株)
研究員



箱石 健太 (はこいし けんた)
日本工営(株)
研究員



一言 正之 (ひとこと まさゆき)
日本工営(株)
課長

