

RAGを用いたヒューマンエラー事象の対策立案

Estimating Preventive Measures for Human Error Events using RAG

ヒヤリハットの集合知を活用

我々は電力会社のヒューマンエラー事象を記録した「ヒヤリハット」の事例を知識源として、ヒューマンエラーに関する対策を立案する手法の検討を行っている。本稿では、大規模言語モデル(LLM)とRAGの枠組みを用い、不完全な「外部知識」をLLMの「内部知識」により補完する拡張方式を提案する。ヒヤリハットの数十万件の大規模事例を用いた対策立案(マルチドキュメント要約)の評価実験では、提案する知識補完の有効性が示唆される結果が得られた。



執筆者
先端技術応用研究所
情報技術グループ
瀬川 修

1 背景と目的

ヒューマンエラーに関する対策立案において、LLM(Large Language Model)をはじめとする生成モデルの適用が考えられる。蓄積された大規模事例から対策を集約し「集合知」として活用できれば、災害防止に大変有用性が高い。

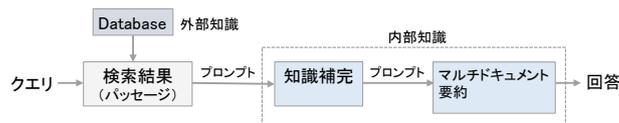
LLMによる知識処理においては、学習に用いていない「外部知識」をプロンプトに加えることにより、ドメイン依存の知識拡張を行うRAG(Retrieval Augmented Generation)と呼ばれる方式が提案され、様々な情報検索や情報推薦に関するタスクに活用されている。RAGの枠組みにおいては、あらかじめ用意した外部知識のデータベースから質問(クエリ)に適合する検索結果(パッセージ)を抽出し、当該パッセージと拡張したクエリでプロンプトを構成し、LLMによって適切な回答を生成する方策が一般的である。しかしながら、外部知識が十分な情報量を有していない場合は、クエリに適合する回答の生成が困難になることが予想される。

これまで外部知識の拡張方式として、クエリの回答に関連する「仮想的な文書」をLLMによって生成し、当該文書に類似した文書群を何らかの知識源のデータベースから検索して「補完知識」として用いるHyDE(Hypothetical Document Embeddings)と呼ばれる手法が提案されている。HyDEではLLMの「内部知識」に基づき外部知識の拡張を行っているが、そもそも外部の知識源に回答に関連する情報が無い場合は、検索タスクを完遂できない。

これに対し、本稿ではヒューマンエラー事象の外部知識に対策が含まれない場合において、LLMの内部知識によって知識補完を行う方策を検討し、拡張された知識を用いて適切な対策をマルチドキュメント要約としてまとめる手法を提案する。

2 提案手法の概要

本稿で提案するRAGの拡張方式の概要を第1図に示す。検索タスクによっては外部知識に必要な情報が含まれていないケースがあり、このような場合、何らかの手段で「知識補完」を行う必要がある。提案手法では検索された複数パッセージに対し、LLMの内部知識によって知識補完(対



第1図 RAGの拡張方式の概要

策生成)を行い、検索結果のコンテキストに加え生成された情報を用いて、マルチドキュメント要約を行う。

(1) 知識検索

従来のRAGと同様に、クエリに適合する検索結果を外部知識のデータベースから得る。文書検索は様々な手法が提案されているが、本稿では潜在意味空間におけるベクトル表現の類似度検索を用いた。

(2) 知識補完

ここで述べる「知識補完」とは、外部知識として不完全な事例(対策が記述されていない場合)に対し情報補完を行うことである。上記で検索された外部知識の個々のパッセージを入力として、LLM(指示チューニングモデル)による対策文の生成を行う。ここでは、外部知識の個々の事例(「発生概要」+「発生状況」)に基づき、プロンプトの指示によって適切な「対策」を推定するものとする。LLMの内部知識を用いることによって、事例のコンテキストから、ある程度のレベルの対策が生成されることが期待される。

(3) マルチドキュメント要約

上記知識補完により得られたパッセージを入力として、LLM(指示チューニングモデル)によるマルチドキュメント要約を行う。対策立案のタスクにおいては、複数事例から対策を要約としてまとめる指示をプロンプトで与えることによって、クエリに応じた対策の「集合知」が生成されることが期待できる。

3 評価実験

(1) 評価データ

電力会社の各部門で蓄積されたヒューマンエラー事象(現場作業、オフィス業務、通勤途上など)を自由記述した「ヒヤリハット」と呼ばれる大規模コーパス(「発生概要」、「発生状況」、「対策」の各レコードが含まれる約35万件の事例、収集期間2019年4月~2024年4月)を用いた。

評価セットとして、対策が記述されている完全知識のセット（「発生概要」、「発生状況」、「対策」）と、対策が記述されていない不完全知識のセット（「発生概要」、「発生状況」）を用意した。評価セット作成においては、情報量のコントロールのため、「発生概要」+「発生状況」は両方で共通のレコードを用い、「対策」の有無のみの違いとする。

(2) 実験条件

今回の評価実験では、電力業務に関連した以下の10個のクエリを用い、後述の各種評価で使用する。

配電の高所作業、配電の活線作業、配電設備巡視、変電所巡視、送電線巡視、水力設備巡視、伐採作業、現場での重量物運搬、熱中症、ドローン

知識検索においてはEmbeddingモデルとして、Multilingual E5を用いた。また、LLMとして、Llama3.1を日本語コーパスで継続学習した指示チューニングモデルであるLlama3.1-8B-EZO-1.1-it（パラメータ数約80億）を用いた。

(3) LLMによる対策生成の評価

「対策」が無い不完全知識を補完する方法として、LLMを用いた対策生成の評価を行った。実験では、前述の10個のクエリで検索された結果の上位30件の各事例の「発生状況」と「発生概要」をLLMに入力し、「対策」を生成する。そして、これらを人間系作成の対策と比較することによって、生成結果の妥当性を評価した。LLMに与えるプロンプトとして「あなたは誠実で優秀なプロの編集者です。以下の事象から対策を考えてください。」を用いた。評価指標としては正解（GT）とのキーワードの一致率を評価するROUGE-1（R1）、ROUGE-2（R2）、ROUGE-L（RL）、およびBERTによる潜在意味空間における類似度を評価するBERT-Scoreを用いた。生成された300件の「対策」の評価結果を第1表に示す。生成結果を目視確認したところ、概ね妥当な対策が生成されていることを確認した。

第1表 LLMによる対策生成の評価

R1	R2	RL	BERT-Score
0.179	0.026	0.138	0.649

(4) 知識補完の有効性評価

提案手法の有効性を確認するため、完全知識の場合と不完全知識の場合における知識補完の有無による条件下でマルチドキュメント要約の比較評価を行った。実験では完全知識を使うケース①と、知識補完を行う3つのケース②～④を比較する。

- ① 「発生概要」 + 「発生状況」 + 「対策」
- ② 「発生概要」 + 「発生状況」 + 「対策（生成）」
- ③ 「対策（生成）」のみ
- ④ 「発生概要」 + 「発生状況」のみ

評価におけるGTは、前述の10個のクエリごとにサンプリングした複数事例30件について、人間系で対策をまとめた要約結果を用いた。生成結果の評価はマルチドクメ

ント要約によって行い、評価指標は前記(3)と同様にR1、R2、RL、BERT-Scoreを用いた。LLMに与えるプロンプトは、前記(3)の対策生成の文言に加え、「以下の複数事例を簡潔に箇条書きで要約してください。要約結果のみを出力してください。」を用いた。評価結果を第2表に示す。また、提案手法②による対策生成の結果例を第3表に示す。

第2表 知識補完の有効性評価

要約の元になる情報源	R1	R2	RL	BERT-Score
① 「発生概要」+「発生状況」+「対策」	0.294	0.065	0.162	0.708
② 「発生概要」+「発生状況」+「対策（生成）」	0.328	0.067	0.170	0.703
③ 「対策（生成）」のみ	0.275	0.055	0.150	0.707
④ 「発生概要」+「発生状況」	0.178	0.029	0.113	0.648

第3表 提案手法②による対策生成例: クエリ「熱中症」

1. 作業時間調整: 熱中症リスクが高まる時間帯（午前10時から午後3時頃）に、作業時間を調整して水分補給と休憩を促進する。
2. 予防体制の強化: 熱中症のリスクが高まる期間に、特に労働者が多い部門や作業場所で早期に対策を講じる。
3. 気温と湿度の監視: 気温が高くなると、水分摂取不足により熱中症が発症するため、温度と湿度を監視し、適切な水分補給を促す。
4. 定期的な水分・塩分補給: 作業員に定期的に水分と塩分を補給させ、実践を確認する。
5. 気象情報のチェック: 休暇前や出向前、現場の気象予報を確認し、日中の高温となる可能性がある場合に十分な対策を講じる。

(5) 考察

第2表のROUGEスコアに着目すると、「生成した対策」を加えた②のケースでは、完全知識の①を上回る結果が得られた。これはLLMで知識補完した情報が、人間系作成の対策に匹敵する情報量を含んでいると考えられ、提案手法の有効性が確認できた。また、②と④のケースの比較では、不完全知識（「発生概要」と「発生状況」）の集合から一足飛びに対策の要約を生成するよりも、一旦それぞれの「対策」を生成して情報源として用いた方が、良好な結果が得られることが分かった。これは段階的に推論を行うことによって、ある種のChain-of-Thoughtのような効果が表れていると考えられる。また、生成した対策のみを用いる③ケースは、①と②の結果に及ばず、要約生成において、「発生概要」と「発生状況」のコンテキストが重要であることを示唆している。

4 まとめ

本稿では、RAGを用いたヒューマンエラー事象の対策立案の一手法として、外部知識の補完によって拡張された知識源を用い、複数事例の対策を集合知としてまとめる手法を提案した。

（本研究は株中電シーティーアイとの共同研究として実施した）