P05

ハルシネーションを抑止する 知識グラフを用いた質疑応答手法の構築

東京科学大学工学院　米持湧　XU Ziwei　市瀬龍太郎

どんな研究？

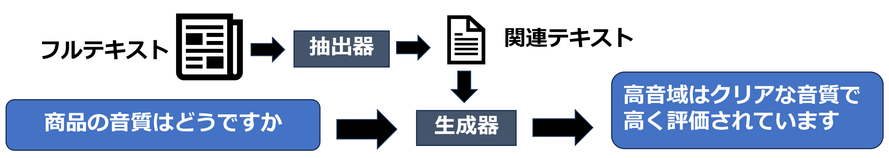
楽天市場の商品に対し，レビューに基づいた評価レポートを生成するための，テキストRAGと知識グラフRAGを統合した手法を提案する．本手法では両RAGを二段階方式で連携し，ハルシネーション検知システムを導入した．検証では，商品情報とレビューから質疑応答データセットと知識グラフを構築した．単語および意味レベルでの類似度評価により，ハルシネーションの抑止と高い信頼性が示された．

研究背景・目的

企業は，自社製品に対する消費者の価値認識や印象を把握するため，アンケート調査を実施している.しかしテキスト形式のアンケート分析は，多数のレビュー読解や課題抽出に多くの時間を要する．　そこで高い分析精度の担保と，要するコストの削減が課題である．そこで本研究では,知識グラフを用いたRAGとハルシネーション検知手法により,低コストで信頼性の高い評価レポートの生成を目指す．

テキストRAGの利点と課題

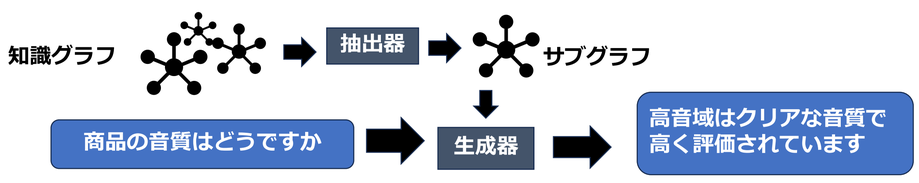
RAG：モデル内の知識と外部知識を統合して回答を生成
　抽出器：質問に関連する外部知識を検索
　生成器：抽出した知識に基づいて応答を生成



テキスト形式の外部知識ではハルシネーションが発生しやすい

知識グラフRAGの利点と課題

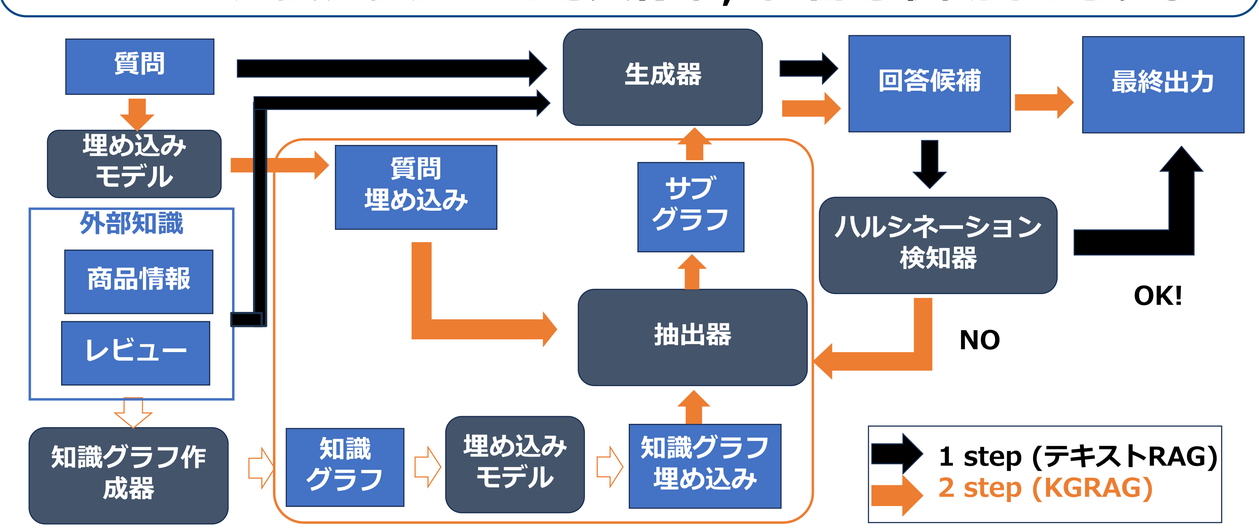
知識グラフを外部知識として活用すると，エンティティ間の関係を直接追跡でき，ハルシネーションの発生を抑止できる



知識が構造的であるため回答は正確，未知の情報には「わからない」と判断可能だが，曖昧な情報や主観的表現を扱いにくい

H-FRAG

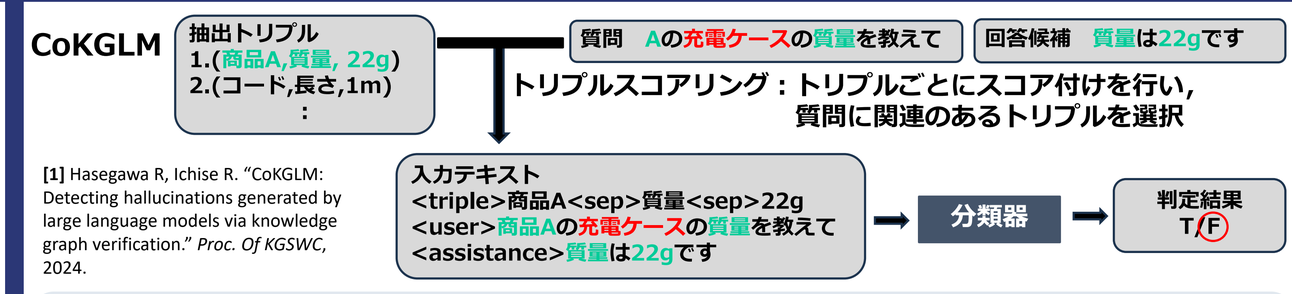
1step：外部知識からテキストRAGを実行し回答候補を出力
ハルシネーション検知器：回答候補がハルシネーションかどうか判別（正しい場合→最終出力，間違いの場合→2step）
2step：・外部知識から知識グラフを構築
　　・知識グラフRAGを実行し，回答を最終出力とする



・ハルシネーション検知器により両手法を整合性を保ち統合
・1stepで曖昧，主観的な知識を必要とする質問に対応
・2段階生成により正確で信頼性の高い回答が可能

ハルシネーション検知器

CoKGLM



[1] Hasegawa R, Ichise R. "CoKGLM: Detecting hallucinations generated by large language models via knowledge graph verification." *Proc. Of KGSWC*, 2024.

知識グラフを利用した検知モデル（CoKGLM）とゼロショットモデル（LLM）をハルシネーション検知器に設定

データセット構築

Q1: レビュー　Q2: 商品情報
Q3: 両方　　Q4: 回答不可

商品情報:[商品,特徴, スペック]
レビュー:[特徴,評価項目, 評価]

楽天市場データから質疑応答データおよび商品知識グラフを作成

推論結果

	ROUGE-1 Total (Q1,Q2,Q3,Q4)	ROUGE-2	ROUGE-L	意味的類似度	時間 (s)
テキスト RAG	0.46 (0.43,0.64,0.42,0.34)	0.33 (0.28,0.53,0.26,0.26)	0.43 (0.40,0.62,0.37,0.34)	0.81	1.15
H-FRAG (LLM)	0.50 (0.41,0.58,0.39,0.61)	0.38 (0.26,0.47,0.24,0.55)	0.48 (0.39,0.57,0.36,0.61)	0.82	2.59
H-FRAG (CoKGLM)	0.32 (0.28,0.25,0.29,0.44)	0.21 (0.15,0.16,0.14,0.37)	0.30 (0.26,0.24,0.25,0.44)	0.75	1.45

質問	正解データ	H-FRAG 1step	H-FRAG(LLM) 2step	H-FRAG(CoKGLM) 2step
この商品は中国語の説明書があるのですか	中国語だと記載されています	わかりません	わかりません	説明書は中国語で提供されます

・ゼロショットベースのH-FRAGが最高精度となる
・1stepの誤答が2stepで改善されている

結論

・両RAGとハルシネーション検知を統合したH-FRAGを提案
・商品情報とレビューからQAデータと知識グラフを構築
・ハルシネーションの抑止と高い信頼性を達成

謝辞

本研究では, 国立情報学研究所のIDRデータセット提供サービスにより, 楽天グループ株式会社の楽天市場データを利用させていただきました. ここに厚く御礼申し上げます.

Institute of SCIENCE TOKYO

連絡先：東京科学大学工学院　米持湧　市瀬龍太郎
Email：yonemochi.y.7760.ac@m.titech.ac.jp, ichise@iee.e.titech.ac.jp
URL：https://www.ai.iee.e.titech.ac.jp/index_j.html