

LLM × KB サーベイ (NLP2023)

2023/05/13

理研AIP 言語情報アクセス技術チーム 研究パートタイマー

林 欽樹

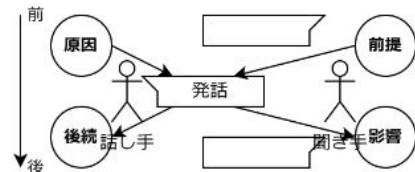
NLP2023 概観

- 知識グラフを利用する研究が盛ん
 - LINE、早大は常識知識グラフ (CSKG) の人手+LLMによる構築に力を入れている
 - 和製chatGPTの開発着手を2023/05/10に発表！(LINEとソフトバンク)
 - 対話
 - 応答生成時に関連エンティティ+**エンティティの要約**を入力する研究
 - 応答生成時に用いるエンティティを **記号化して知識の使い方**に着目した学習を行う研究
 - GNN
 - GNNで構築された文献グラフに関する研究
 - 外部ツールとの連携
 - 複数の外部ツールを用いた Chain of Thought手法で数値計算タスクを解く研究
 - KG × BERT
 - KBから抽出した埋め込み表現を BERTの埋め込み表現に連結してタスクを解く研究

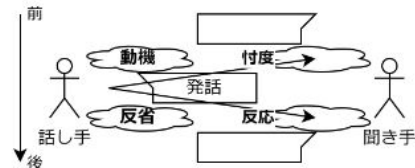
対話に基づく常識知識グラフの構築と 対話応答生成に対する適用

井手竜也¹ 榮田亮真¹ 河原大輔¹
山崎天² 李聖哲² 新里顕大² 佐藤敏紀²
¹早稲田大学理工学術院 ²LINE 株式会社

- 常識に基づく対話応答生成に向けた対話常識グラフの提案(今後公開予定)
 - カテゴリ(出来事、心情)、時系列(前後)、対象(話し手、聞き手)などの次元を考慮し、推論すべき関係を考慮
 - 英語には複数存在するが、日本語には存在しない
 - 先行研究にはなかった「どの話者に向けた推論か」という次元に注目
- HyperCLOVA JPにおけるIn-Context Learningにおいて心情の推論をプロンプトとして付与
 - 生成される応答の特徴が変化した
- コーパス上の発話に対してクラウドソーシングで推論を付与
- 構築した対話常識グラフに含まれる最後から二番目までの発話を履歴とし、最後の発話を生成
 - 推論を付加する場合としない場合を比較(相手に対する)
- BLEU、distinct、PPLを用いた自動評価においては推論を明示的に与えた方が人間に近い応答を生成することが示された
- 人手評価に関しては、「動機と反省」に関する推論を与えるより、「忖度と反応」に関する推論を与えた方が評価が高かった
 - 人間らしさよりも、相手を慮った回答が評価されている



(a) 出来事



(b) 心情

図2 対話常識グラフで推論すべき関係

鈴木と佐藤の二人が会話している。
鈴木が反対と思っている、と佐藤は考える。そして、佐藤は恥ずかしいと思う。
佐藤「まってwww服裏表反対でご飯食べに来てたwww」
鈴木が早く着替えて来てとかギャグかと思った、と佐藤は考える。そして、佐藤は恥ずかしすぎると思う。
鈴木「ええなんで笑笑」
鈴木がいつから反対だったのかと思っている、と佐藤は考える。そして、佐藤は失敗とか失敗したが、大したミスではないとか錯んで行かないと思う。
佐藤「夕飯行く前に寝てたから服替え直したんだけどその時にミスってwww」
鈴木がおっちょこちょいだとか恥ずかしいと思った、と佐藤は考える。そして、佐藤はアホやろと思う。
鈴木「気づいた時恥ずかしい笑笑」
鈴木が恥ずかしいとかどうしてそうなるのか、と思っている、と佐藤は考える。
そして、佐藤は恥ずかしいと思う。
佐藤「いやほんとはずかしかったwww」

人間と言語モデルに対するプロンプトを用いた ゼロからのイベント常識知識グラフ構築

井手竜也¹ 村田栄樹¹ 堀尾海斗¹ 河原大輔¹
 山崎天² 李聖哲² 新里顕大² 佐藤敏紀²
¹早稲田大学理工学術院 ²LINE株式会社

- クラウドソーシングと大規模言語モデルを併用して知識グラフを構築する手法を提案
 - 知識の記述に関して、少しの例から多くの例を作成してもらう、という意味で両者は本質的に同じ
- クラウドソーシングによって小規模な知識グラフを構築し、それをプロンプトに用いることで大規模言語モデルが持つ知識を抽出
 - 時間的、金銭的成本を削減することを期待
- 小規模な知識モデル(GPT-2、T5)の訓練を行い、BLEU、BERTscoreによる自動評価、人手評価を実施
 - 自動評価では、T5よりGPT-2が勝った
 - 人手評価による尤もらしさはおおむね9割を達成

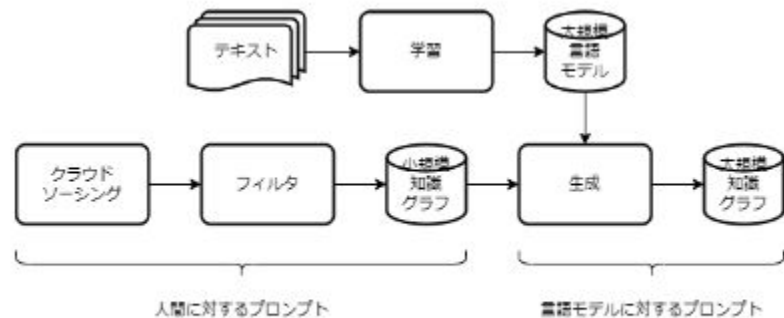


図1 提案手法の概要

表7 GPT-2に基づく知識モデルの生成例

イベント	関係	推論
Xがパソコンで仕事を する	必要	Xがパソコンを起動する
	影響	Xが残業する
	意図	お金を稼ぎたい
	反応	疲れた

大規模言語モデルによって構築された常識知識グラフの 拡大と低コストフィルタリング

村田栄樹¹ 井手竜也¹ 榮田亮真¹ 河原大輔¹
山崎天² 李聖哲² 新里顕大² 佐藤敏紀²
¹早稲田大学理工学術院 ²LINE 株式会社

表2 小規模グラフから採用する訓練データの例. 各例は (h, r, t) を表す.

負例タイプ	もととなる正例	採用する負例
タイプ1	(Xが顔を洗う, xNeed, Xが水を出す)	(Xが水を出す, xNeed, Xが顔を洗う)
タイプ2	(Xが顔を洗う, xNeed, Xが水を出す) (XがPCで仕事をする, xNeed, XがPCを起動する)	(Xが顔を洗う, xNeed, XがPCを起動する)
タイプ3	(Xが顔を洗う, xNeed, Xが水を出す) (Xが顔を洗う, xEffect, Xが顔を拭く)	(Xが顔を洗う, xNeed, Xが顔を拭く)

- 日本語常識知識グラフ(CSKG)における生成に関するファクタの分析及び低コストなフィルタリング手法の提案
- 既存のCSKGをショットとして与え、HyperCLOVAの持つ知識をCSKGに引き出すために必要な小規模グラフのサイズを評価
 - イベント数は250-300程度で十分
 - イベントあたりの推論回数は100程度に増やしてもユニークな推論を行うことができる(既存研究では10程度)
- 低コストフィルタリング
 - フィルタリングには負例が必要
 - 小規模CSKGのノードのheadとtailを意図的に入れ替えたものを擬似的な負例として採用
 - RoBERTa-largeを用いてフィルタモデルを訓練
 - JGLUEに含まれるJNLIデータセットを用いて学習・評価
 - 適切な閾値を設けることで9割程度の適切なフィルタリングができた
- 簡易的な常識モデルの訓練・評価
 - GPT-2をCSKGで訓練
 - データサイズが小さい時はフィルタが有効
 - 適切なデータと訓練データのサイズにトレードオフがある?
 - 大きい時は変化なし

知識グラフと Wikipedia を用いた雑談対話モデルの構築

郭 恩孚¹ 南 泰浩¹

¹ 電気通信大学情報理工学研究所

- 雑談用の対話モデルにおいて、ユーザーが提示したエンティティに関して対話を広げるために、対話履歴から次の話題となるエンティティとその要約を外部から取得し、モデルの入力文脈に追加する Fine-tune手法を提案
 - 先行研究では外部知識トリプルのみを利用していたが、過去の対話履歴を持ち出せないこと、新たな話題を提示できないことなどが問題だった
- Wikidataから構築した知識グラフから次の話題となるエンティティが存在するか確認し、次のエンティティの要約を入力文脈に追加
 - 事前学習はWikiコーパス、CC-100コーパス、ツイートトリプライコーパス
 - Fine-tuneはツイート擬似対話コーパス
- 5名の評価者に対して知識トリプルのみを用いたベースラインモデル及び提案モデルが生成した対話を提示し、自然性、妥当性、適切性について 5段階評価
 - 自然性は0.66、妥当性が0.52、適切性が0.6向上
- ライオン(企業)とライオン(動物)を紐付けるなどの飛躍もあった
 - 知識グラフの適切性が求められる

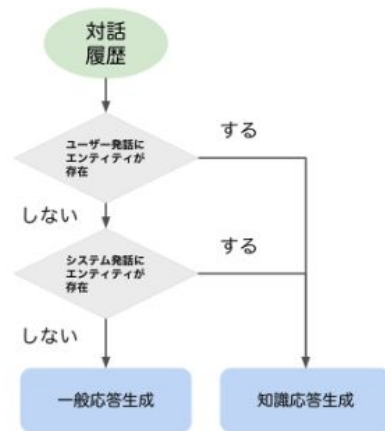


図3 処理のフローチャート

知識グラフに基づく応答文生成における 記号化されたエンティティの利用

梶井裕貴¹ 滝口哲也² 有木康雄²

¹ 神戸大学工学部情報知能工学科 ² 神戸大学大学院システム情報学研究科



図2 エンティティ記号化を利用した応答文生成

- 雑談対話モデルの応答文生成に関して、具体的な知識を学習するより、具体名を記号化して、知識の使い方を抽象化して学習した方が良いという仮説を検証
 - 既存手法よりも多くの知識を応答文に取り入れることが可能であることを確認
- 既存手法のEntity Name Constrained Decoding(ENCD)では応答文のどこに知識を含めるべきか学習を行う
 - エンティティを"[]"で囲うことで参照タイミングを学習(エンティティタグ付き学習)
 - このまま用いると具体的な名称を学習に含めてしまいパラメータの無駄遣いになってしまうので、提案手法ではエンティティを記号化する
 - 外部から知識を得るのであれば、学習時にエンティティの具体的情報はいらぬという仮説
 - エンティティを SKA、SKB などの特殊トークンに置換して学習
- F1、KF1を用いて3通り(ベースライン、ENCD、提案手法)のFine-tuneをした T5を比較
 - 提案手法が既存手法を上回った
 - 従来手法より「知識の使い方」に特化した学習ができています

テキスト情報の表現を利用した文献グラフの表現学習

片桐 脩那 井田 龍希 三輪 誠 佐々木 裕
豊田工業大学

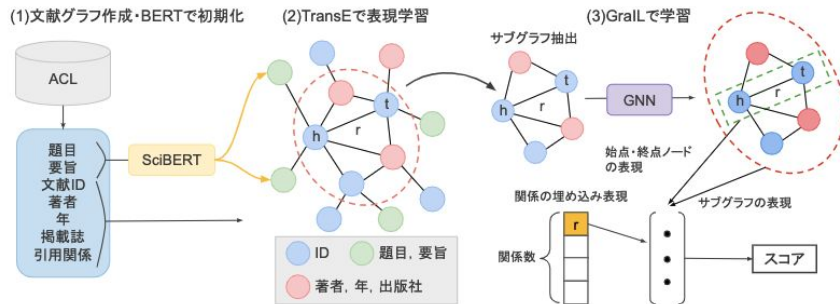


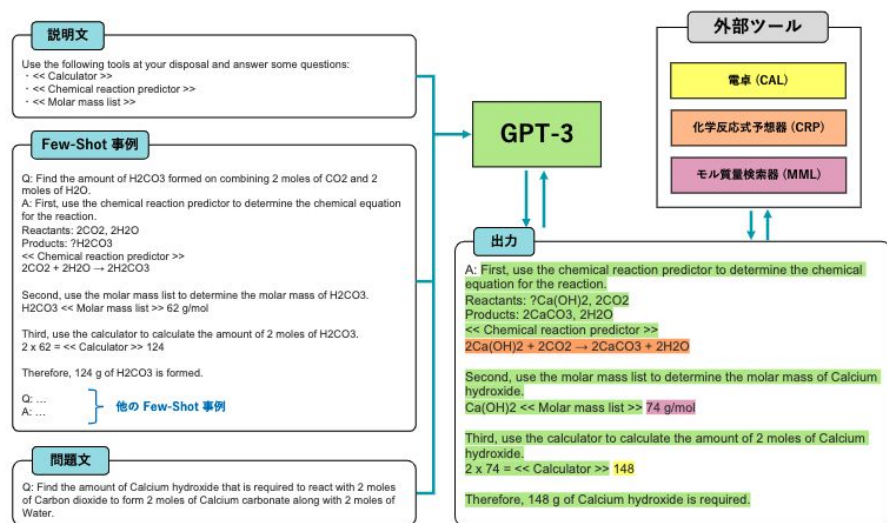
図1 文献グラフの作成と表現学習の流れ

- 論文の文献情報を表す文献グラフの構築における、グラフ埋め込み、GNNsを用いたリンク予測手法の利用可能性の評価
- 先行研究ではグラフ埋め込み、GNNsを用いた手法は十分に評価できていない
 - 題目や要旨などのテキスト情報に SciBERTによって埋め込み表現を得て、TransEで表現学習
 - リンク予測には GraILを利用してGNNの学習を行う
- ACL Anthologyの過去5年の文献情報(16,916件)を学習に使用
- 3つのモデルを比較し、結果は GraIL < TransE < TransE+BERT
 - 文献IDのノードと著者の関係にあるノードを予測するリンク予測タスクで評価
 - TransE: 初期表現にランダムな表現を使用
 - TransE+BERT: SciBERTを用いた埋め込みを初期表現として使用
 - GraIL: TransE+BERT
- GNNsが低評価になる原因の考察
 - グラフをスター型で構築したために到達するためのホップ数が大きくなっている
 - 出版年など頻出する情報が悪影響を及ぼしている
- グラフの構築方法に改善の余地あり

大規模言語モデルに基づく 複数の外部ツールを利用した推論フレームワーク

稲葉 達郎 清丸 寛一 Fei Cheng 黒橋 禎夫
京都大学

- LLMによる推論のさらなる高度化を目指し、検索機や電卓などの複数の外部ツールの利用を可能にするフレームワークの提案
 - 推論過程を含めて回答を生成するようにプロンプトを設計し、推論過程の一部として外部ツールを呼び出す文字列を生成させる
 - 実行時に外部ツールの呼び出しが発生したら、推論過程を中断し、外部ツールの実行結果を推論過程に追記して推論過程を再開
- プロンプトは利用可能な外部ツールを明示した **説明文**、いくつかの問題文と解答のペア (**Few-Shot 事例**)、解かせる **問題文** を結合して構成
- NumGLUEのTask 2(化学の専門知識を要する数値推論)を以下のツールを用いて解く
 - Calculator: Pythonのeval function
 - Chemical reaction predictor: 反応物と生成物の入力から化学反応式を出力
 - Molar mass list: 化学式を入力としてその守る質量を出力
- (Zero-shot | Few-shot) × (None | CoT) × (外部ツールの組み合わせの中から6手法を比較検討
 - 提案手法がSOTA



日本語大規模言語モデルにおける 知識グラフを活用した意味理解性能の向上

中本裕大¹ 瀬在恭介¹ 元川凱喜¹ 麻生英樹² 岡崎直観³

¹SCSK 株式会社 ²産業技術総合研究所 ³東京工業大学

表 2: 実験結果

Model	MARC-ja	JSTS	JNLI	JSQuAD	JCommonsenseQA	子育て FAQ
	acc	Pearson/Spearman	acc	EM/F1	acc	acc(平均)
BERT	95.31	87.06 / 81.91	84.56	82.67 / 90.75	75.16	91.66
提案手法	95.68	87.58 / 82.70	85.88	81.97 / 90.13	75.72	92.91

- BERTにドメイン固有の知識を注入するために、KG (Wikidata)を入力に結合してFine-tuneする手法の検討
 - 英語、中国語では盛んに行われているが、日本語では未検討
- 入力文についてMeCabによって分かち書きを行い、名詞のうちKGでヒットするものを入力に追加する
 - KGはTransEによってWikidataの埋め込み表現を獲得しておく
- JGLUE、子育てFAQデータによって学習・評価
 - ほとんどのタスクで性能が向上
 - LIMEによる属性の貢献度分析
 - 知識があることで新たに正解・不正解になる例が存在する
 - 文脈・語義に沿った知識注入のためにエンティティリンキングを正しく行う必要性がある

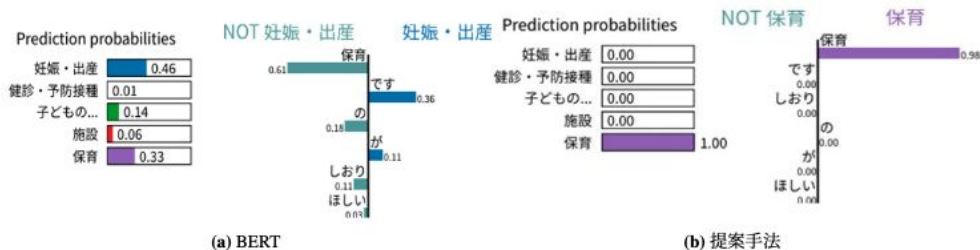


図 2: 提案手法で新規に正解した例

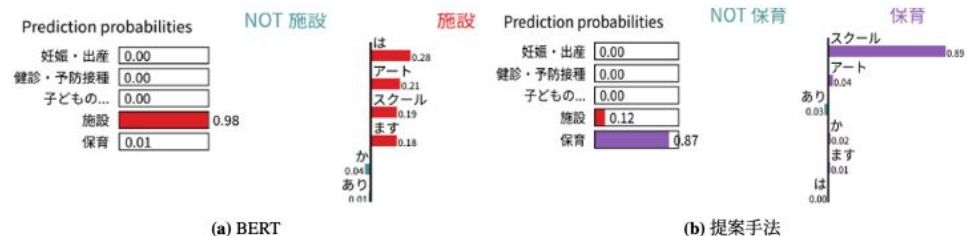


図 3: 提案手法で新規に不正解になった例