

令和 5 年 5 月 29 日現在

機関番号：14603

研究種目：若手研究

研究期間：2021～2022

課題番号：21K17801

研究課題名（和文）共参照クラスタを明示的に推定する先行詞の解析誤りに対し頑健な共参照解析手法

研究課題名（英文）Towards Robust Coreference Resolution by using Global Information of Entities

研究代表者

上垣外 英剛（Kamigaito, Hidetaka）

奈良先端科学技術大学院大学・先端科学技術研究科・准教授

研究者番号：40817649

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：近年登場した、事前学習済み言語モデルを用いて大域的な情報を活用可能な共参照解析器を知識グラフにより支援することを目的に、知識グラフに対する補完を学習する際の理論的な分析を行なった。また、その分析に基づいて、疎な知識グラフにおける補完性能を向上させる技術についても提案した。さらに事前学習済み言語モデルが言語と画像の融合分野へと進出したことに対応し、言語と画像の両面からモデルに含まれるエンティティに関する知識を検証するタスク及びデータセットを作成し、またどのような構造のモデルが適しているかについても検証した。さらにエンティティを利用して自然言語生成モデルを制御可能かについても検証した。

研究成果の学術的意義や社会的意義

本研究では事前学習済み言語モデルに基づいた共参照解析手法を支援することを目的に、知識グラフの補完性能の向上及び視覚と画像の融合分野において事前学習済み言語モデルに含まれている知識についてを検証するためのタスク及びデータセットを提案し、実際の検証を行なった。これらの成果は事前学習済み言語モデルの学習過程では通常明示的に与えられないことがない知識を扱うものであり、昨今の事前学習済み言語モデルの急速な発展の後にも利用価値を持つことが期待できる。特に視覚と画像の融合分野においてはテキストの生成および画像の生成の両面において性能向上の観点から重要な研究成果であり、共参照解析に限定されない活用が期待される。

研究成果の概要（英文）：To support the recently appeared coreference analyzers that can utilize global information using pre-trained language models by using knowledge graphs, we conducted a theoretical analysis of learning completions for knowledge graphs. Based on the analysis, we also proposed techniques to improve completion performance on sparse knowledge graphs. In response to the expansion of pre-trained language models into the area of vision and language fields, we created a task and a dataset to verify the knowledge of entities in the models from both vision and language perspectives, and also to verify which model structure is suitable for the task. We also tested whether entities can be used to control natural language generation models.

研究分野：自然言語処理

キーワード：知識グラフ 知識グラフの補完 タスク 知識グラフの埋め込み 負例サンプリング損失 視覚と言語の融合タスク 視覚と言語の融合モデル 事前学習済み言語モデル BERT

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

研究開始当初、自然言語処理分野では共参照解析を行うための手法として、予測されたそれぞれの言及に対応する先行詞を逐次的に独立に予測し、最終的に予測された言及と先行詞との対応関係を抽出することで共参照解析を行うという手法 [1] が主流であった。この解析手法は ELMo や BERT といった事前学習済みモデルの登場により性能が大きく向上し、また SpanBERT や RoBERTa といった BERT の改良型の前学習済みモデルの出現によりさらなる性能向上が達成されている段階にあった。その一方で、共参照解析を行うための手法そのものは事前学習済み言語モデルの成功以前からほぼ同様のものが使用されており、事前学習済み言語モデルに含まれる知識を活用しきれていない状況にあった。

[1] Lee, K., He, L., Lewis, M., & Zettlemoyer, L. “End-to-end Neural Coreference Resolution” In Proc. of EMNLP2017.

2. 研究の目的

本研究提案の目的は共参照解析において各言及に対する先行詞の予測を逐次的に行うことで生じる頑健性の低下を、共参照クラスタを予測時に直接扱うことで対処しようというものであった。直接共参照クラスタを予測することにより、事前学習済みモデルが持つ知識を予測に反映可能となることを期待できる。しかし、事前学習済みのモデルを用いたほぼ同様の手法 [1] が提案されたため、事前学習済み言語モデルの進展の後にも利用可能な共参照解析の補助手段として知識グラフに着目し、知識グラフの補完の性能向上及び事前学習済み言語モデルが持つ知識の調査を目的とした。

[1] Wu, W., Wang, F., Yuan, A., Wu, F., & Li, J. “CorefQA: Coreference resolution as query-based span prediction”, In Proc. of ACL2020.

3. 研究の方法

知識グラフの補完に関しては、利用されている損失関数に関する理論的な分析を行なった。そして、その理論的な分析の結果に基づき、補完性能を向上させる手法の開発を実施した。さらに事前学習済み言語モデルの言語と画像の融合分野での進展も考慮し、事前学習済み言語モデルに含まれる知識を調査する上で有用な言語と画像の双方を含むデータセットを作成し、実際に検証を行なった。

4. 研究成果

知識グラフの補完で頻繁に利用されている損失関数を様々なスコアリング法と組み合わせて使用する際の適切な利用法について理論的に明らかにした。さらにその理論的な検証に基づき、疎な知識グラフでの学習を扱う際に有用な訓練データにおける知識グラフの枝の出現頻度を調整するサブサンプリング手法を提案しその有用性を実験により示した。本研究内容を言語処理学会第 28 回年次大会にて発表を行い、その内容で優秀賞を受賞した [1]。また同内容により詳細な実験を加えたものが機械学習分野のトップ国際会議の一つである ICML に採録された [2]。また頻度を利用したサブサンプリング手法はサブサンプリングそのものが枝の出現頻度に大きく影響されるため、モデルによる出現頻度を用いる新たなサブサンプリング法についても提案し、同様にその有用性を実験により示した [3]。さらに視覚と言語の融合分野で使用されている事前学習済みモデルに含まれるエンティティに関する知識を検証するためのタスク及びデータセットを提案し、既存のモデルが学習過程でエンティティに関する知識を忘却していることを示した。この研究内容は言語処理学会第 29 回年次大会にて委員特別賞を受賞した [4]。さらに、同内容に詳細な実験を加えた研究が自然言語処理分野のトップ国際会議の一つである ACL2023 に採録された [5]。さらに視覚と言語の融合分野で使用されている事前学習済みモデルに対して言語情報を保持する上で適したモデルのネットワーク構造についても検証した [6]。そして言語モデルが自然言語生成を行う上でのエンティティによる制御可能性についての検証も実施し、強化学習による制御が可能であることを示した [7]。

- [1] 上垣外英剛, 林克彦. 2022. “知識グラフ埋め込みにおける負例サンプリング損失の分析”, 言語処理学会第 28 回年次大会 (NLP 2022). (優秀賞)
- [2] Hidetaka Kamigaito, Katsuhiko Hayashi. 2022. “Comprehensive Analysis of Negative Sampling in Knowledge Graph Representation Learning”, In Proc. of ICML 2022.

- [3] Xincan Feng, 上垣外英剛, 林克彦, 渡辺太郎. 2023. "知識グラフ補完のためのモデル予測に基づくサブサンプリング", 言語処理学会第 29 回年次大会 (NLP 2023).
- [4] 上垣外英剛, 林克彦, 渡辺太郎. 2023. "視覚と言語の融合モデルにおける知識の振る舞いを調査するための表と画像の生成タスクの提案及びその調査結果", 言語処理学会第 29 回年次大会 (NLP 2023). (委員特別賞)
- [5] Hidetaka Kamigaito, Katsuhiko Hayashi, Taro Watanabe. 2023. "Table and Image Generation for Investigating Knowledge of Entities in Pre-trained Vision and Language Models", In Proc. of ACL2023.
- [6] 白井尚登, 上垣外英剛, 渡辺太郎. 2023. "エッジプロベリングを用いた事前学習済みの視覚と言語に基づくモデルにおける言語知識の分析", 言語処理学会第 29 回年次大会 (NLP 2023).
- [7] 星野智紀, 上垣外英剛, 渡辺太郎. 2023. "忠実性向上のために n-gram の抽出性を報酬とする強化学習を用いる抽象型要約", 言語処理学会第 29 回年次大会 (NLP 2023).

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 上垣外英剛, 林克彦
2. 発表標題 知識グラフ埋め込みにおける負例サンプリング損失の分析
3. 学会等名 言語処理学会第28回年次大会 (NLP 2022)
4. 発表年 2022年

1. 発表者名 Hidetaka Kamigaito, Katsuhiko Hayashi
2. 発表標題 omprehensive Analysis of Negative Sampling in Knowledge Graph Representation Learning
3. 学会等名 The 39th International Conference on Machine Learning (ICML 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Xincan Feng, 上垣外英剛, 林克彦, 渡辺太郎
2. 発表標題 知識グラフ補完のためのモデル予測に基づくサブサンプリング
3. 学会等名 言語処理学会第29回年次大会 (NLP 2023)
4. 発表年 2023年

1. 発表者名 星野智紀, 上垣外英剛, 渡辺太郎
2. 発表標題 忠実性向上のためにn-gramの抽出性を報酬とする強化学習を用いる抽象型要約
3. 学会等名 言語処理学会第29回年次大会 (NLP 2023)
4. 発表年 2023年

1. 発表者名 白井尚登, 上垣外英剛, 渡辺太郎
2. 発表標題 エッジローピングを用いた事前学習済みの視覚と言語に基づくモデルにおける言語知識の分析
3. 学会等名 言語処理学会第29回年次大会 (NLP 2023)
4. 発表年 2023年

1. 発表者名 上垣外英剛, 林克彦, 渡辺太郎
2. 発表標題 視覚と言語の融合モデルにおける知識の振る舞いを調査するための表と画像の生成タスクの提案及びその調査結果
3. 学会等名 言語処理学会第29回年次大会 (NLP 2023)
4. 発表年 2023年

1. 発表者名 Hidetaka Kamigaito, Katsuhiko Hayashi, Taro Watanabe
2. 発表標題 Table and Image Generation for Investigating Knowledge of Entities in Pre-trained Vision and Language Models
3. 学会等名 The 61st Annual Meeting of the Association for Computational Linguistics (ACL2023)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>言語処理学会第28回年次大会表彰一覧 (NLP2022) https://www.anlp.jp/nlp2022/award.html 言語処理学会第29回年次大会表彰一覧 (NLP2023) https://www.anlp.jp/nlp2023/award.html</p>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------