

様 式 F - 7 - 1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（令和２年度）

			機関番号	1 4 6 0 3
所属研究機関名称		奈良先端科学技術大学院大学		
研究 代表者	部局	先端科学技術研究科		
	職	教授		
	氏名	渡辺 太郎		

1．研究種目名 研究活動スタート支援 2．課題番号 20K23325

3．研究課題名 知識グラフを統合したニューラル機械翻訳

4．補助事業期間 令和２年度～令和３年度

5．研究実績の概要

深層学習により大幅な性能向上を果たしたニューラル機械翻訳は、モデルの学習のために、大量のデータを必要とする。ところが、データを増やすだけでは、固有表現や、誕生日などの属性、所属先など他のオブジェクトとの関連性など、日々更新される知識を翻訳するのは難しい。本研究では、物事の属性および関連性を記述し、かつ、不完全ながらも多言語化された知識グラフを統合した機械翻訳を実現することで、問題が解決できるかを解明する。

2つの手法を提案した。一つは、原言語を固有表現抽出器で固有表現を同定し、その位置を特殊な記号へ変換、また、その特殊な記号に対するベクトル表現を予め学習された知識グラフのベクトル表現を用いる手法である（KG-tag NMT）。もう一つは、特殊な記号への変換はせず、単語よりもより粒度の細かい、サブワード単位に学習された知識グラフのベクトル表現を用いる手法である（Subworded-KG NMT）。KG-tag NMTでは、固有表現単位にベクトル表現を学習するため、知識グラフの知識を直接反映できるが、原言語の入力文と必ずしも対応する固有表現が存在するとは限らない。Subworded-KG NMTでは、固有表現をサブワード単位へ分割し、エンコーダ・デコーダの枠組みで学習する手法を実現した。このため、固有表現を直接反映したベクトル表現ではないが、カバレッジが非常に大きくなり、入力文のほぼ全ての固有表現に対し、知識グラフを反映可能である。

英独機械翻訳のベンチマークで用いられるWMT14では、従来法と比較して、KG-tag NMTでは性能向上が見られなかったが、Subworded-KG NMTでは、機械翻訳の標準的な評価尺度BLEUで1ポイント以上の向上が見られた。この結果から、提案手法の有効性が示された。

6．キーワード

機械翻訳 知識グラフ

7．現在までの進捗状況

区分 （２）おおむね順調に進展している。

理由

本研究では、知識グラフをニューラル機械翻訳と統合することにより、大量のデータに依存せずとも高精度な機械翻訳を実現でき、かつ、説明可能な翻訳を生成可能であることを明らかにする。このため、複雑なグラフ構造がニューラル機械翻訳へと統合が可能であることを初年度の目標として設定した。

この目標を実現するため、1)知識グラフがニューラル機械翻訳と統合できるよう、知識グラフで記述された物事の属性および関連性を数値ベクトルへと表現する手法を実装する、2)原言語の入力文に対し、多言語知識グラフとの対応付を行い、対応関係が取れた物事に対するベクトル表現を組み合わせることで翻訳を生成するニューラル機械翻訳システムを実装する、といった2つの実装および評価を具体的な目標とした。

これまでの研究により、2つの手法、KG-tag NMTおよびSubworded-KG NMTを提案した。2つの手法は知識グラフを数値ベクトルへと表現し、かつ、原言語の固有表現に対して対応するベクトル表現を組み合わせる、非常に簡単な手法である。実験結果から、機械翻訳の標準的な評価尺度BLEUと比較した場合、従来法より性能向上が見られた。このことから、提案法の有効性が示され、かつ、研究計画どおりに順調に進展しているといえる。

2 版

8．今後の研究の推進方策

現在の手法では、知識グラフの固有表現をサブワード単位にしたモデル(Subworded-KG NMT)では、サブワード間のコンテキストを全く考慮していないものである。複数のサブワードで構成される、単語あるいは単語列で表される固有表現全体のベクトル表現を実現することでさらに性能の向上を目指す。実装したシステムを機械翻訳研究で標準的なベンチマークとして使われるWMT20の日英および中英翻訳タスク(<http://www.statmt.org/wmt20>)にて評価する。学習データが半分や1/4など意図的に少なくした状態であってもDBPediaなどの知識グラフを統合することで、翻訳結果がほぼ同等あるいはそれ以上の性能が達成可能であることを示す。この実験により、提案手法が大量のデータに依存せずとも高精度な機械翻訳を実現できることを明らかにする。

9．次年度使用が生じた理由と使用計画

当該年度で計算機を購入し、本研究のための実験環境を整える予定だったが、情報通信研究機構のご厚意により、無償で計算機を使用可能であったため、必要がなくなった。
未使用額および翌年度分として請求した額は、計算機を購入および国際会議での発表のため支出予定である。

10．研究発表（令和2年度の研究成果）

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件／うち国際学会 0件）

1．発表者名 坂井優介，渡辺太郎，藤田篤
2．発表標題 知識グラフ埋め込みを用いたニューラル機械翻訳におけるエンティティ表現の改良
3．学会等名 言語処理学会第27回年次大会
4．発表年 2021年

〔図書〕 計0件

11．研究成果による産業財産権の出願・取得状況

計0件（うち出願0件／うち取得0件）

12．科研費を使用して開催した国際研究集会

計0件

13．本研究に関連して実施した国際共同研究の実施状況

-

14．備考

-