

令和元年5月30日現在

機関番号：12601

研究種目：若手研究(A)

研究期間：2016～2018

課題番号：16H05872

研究課題名（和文）マルチモーダル深層エンコーダ・デコーダネットワークによるゼロショット機械翻訳

研究課題名（英文）Zero-shot machine translation using multimodal deep encoder-decoder networks

研究代表者

中山 英樹（Nakayama, Hideki）

東京大学・大学院情報理工学系研究科・准教授

研究者番号：00643305

交付決定額（研究期間全体）：（直接経費） 18,000,000円

研究成果の概要（和文）：本研究では、画像を異言語テキスト対応付けのハブとして活用することで、対訳テキストコーパスを必要とせず、ソース言語・ターゲット言語それぞれの画像付き単一言語ドキュメントのみから学習する機械翻訳手法を実現した。さらに、翻訳の出力多様性の向上や、手法の軽量化によるスケーラビリティの向上など、実用性を高める各種の改良を行った。これらの成果はACL、ICLRをはじめとするトップレベル国際会議に複数採択されると共に、言語処理学会年次大会で最優秀賞を二度受賞するなど高い評価を得た。

研究成果の学術的意義や社会的意義

機械翻訳はより一層の技術革新が強く求められているアプリケーションであるが、現在の一般的なアプローチにおいては、学習に用いる対訳テキストコーパスの量が性能向上の鍵となる。しかしながら、実際には同一内容を複数言語で記述したテキストドキュメントは少なく、GAFa等一部の巨大企業にデータを独占されているのが現状である。本研究で提案するアプローチでは、誰でも比較的容易に入手可能な画像付き単一言語ドキュメントのみからの学習を実現するものであり、学術的にも独創的な試みであると同時に、機械翻訳の民主化に貢献しうる点で社会的意義も大きいものであると考える。

研究成果の概要（英文）：In this research, we have developed a zero-shot machine translation method which can be trained only with monolingual image-text data, without the help of parallel text corpus. This method is realized by the idea of using images as a hub to align texts in different languages. Moreover, we have improved the method in many aspects to enhance its practicality such as output diversification and speeding up. These results are accepted at many top-level international conferences such as ACL and ICLR, and awarded the best paper awards twice at the NLP domestic conference.

研究分野：画像認識、自然言語処理

キーワード：機械翻訳 ゼロショット学習 マルチモーダル 画像認識 ニューラルネットワーク 表現学習

1. 研究開始当初の背景

機械翻訳は人工知能の最も重要なアプリケーションの一つである。その実現の鍵となるのは、異言語の翻訳文が対となった対訳テキストコーパスを用いた機械学習であり、**Web** ビッグデータに代表される大規模データによる精度向上が期待されている。しかしながら、実際には同一内容を複数言語で記述したテキストドキュメントは少なく、**Wikipedia・Web** ニュース・国際会議の議事録・特許文書など一部のドメインにしかまとまった量は存在しない。自然言語処理分野における伝統的な機械翻訳研究ではこれらの一部データを実験的に対訳コーパスとして用いてきたが、小規模かつドメイン依存のデータであるため汎用性に乏しい。より一般的な **Web** ドキュメントは単一の言語のみで書かれている場合がほとんどであり、直接対訳コーパスとして用いることは不可能である。すなわち、**Web** ビッグデータ時代にあっても、機械翻訳では現実的にはごく一部のスモールデータしか用いる手立てがないのが現状である。

一方で、一般のブログ・**SNS** 投稿などにおけるテキストは、単一言語に閉じている一方、画像等のリッチなマルチメディアデータを伴うことが一般的になっている。これらの画像はテキスト内容と強い関連性を有することが期待できる。また、画像の内容は母国語に依らず理解できるように、言語によらない普遍的な手がかりとなるモダリティであると言える。最近では深層学習の成功を背景に、画像からの一般的な物体認識の精度が人間並みに到達していることから、画像特徴表現を介して異言語テキスト間の関連性を推定することが十分可能であると考えられる。

以上のような背景から申請者は、画像を異言語テキスト対応付けのハブとして活用することで、一切の管理された対訳テキストコーパスを必要とせず、各言語に閉じたドキュメントのみからゼロショットで機械翻訳システムを学習する手法を提案する。本研究により、従来とは比較にならない規模のビッグデータを機械翻訳へ統合活用する道筋が開けると考える。

2. 研究の目的

本研究では、対訳テキストコーパスを一切用いず、ソース言語・ターゲット言語それぞれの画像付き単一言語ドキュメントのみから機械翻訳システムを学習させることを目的とする。これを実現するためには、異言語テキストを画像を介して間接的に接続するためのメカニズムが必要である。また、性能を上げるためには大規模なデータからの学習に耐えるための手法の高速化・効率化や、出力（翻訳文）の多様性を増すための工夫も重要である。大きく分けて、以下の **3** 項目が本研究で取り組んだ課題である。

- 項目 1：マルチモーダルエンコーダ・デコーダネットワークによるゼロショット翻訳の実現
- 項目 2：デコーダの改良による翻訳多様性の向上
- 項目 3：ネットワークの高速化と軽量化

3. 研究の方法

項目 1：マルチモーダルエンコーダ・デコーダネットワークによるゼロショット翻訳の実現

画像側に畳み込みニューラルネット、テキスト側にリカレントニューラルネットを用い、埋め込みレイヤにおいて統合することで、画像・テキストの共通概念を埋め込むマルチモーダルエンコーダネットワークを構築する。結果として、画像を軸としてソース言語・ターゲット言語の近似的なアラインメントがなされた埋め込み表現が得られる。この埋め込み表現から、リカレントニューラルネットにより目標とする言語の文生成を行うデコーダネットワークを構築することにより、推論時においてはソース言語テキストのみの入力から翻訳文を出力することが可能となる（図 1）。以下、これを手法 1 と呼ぶ。

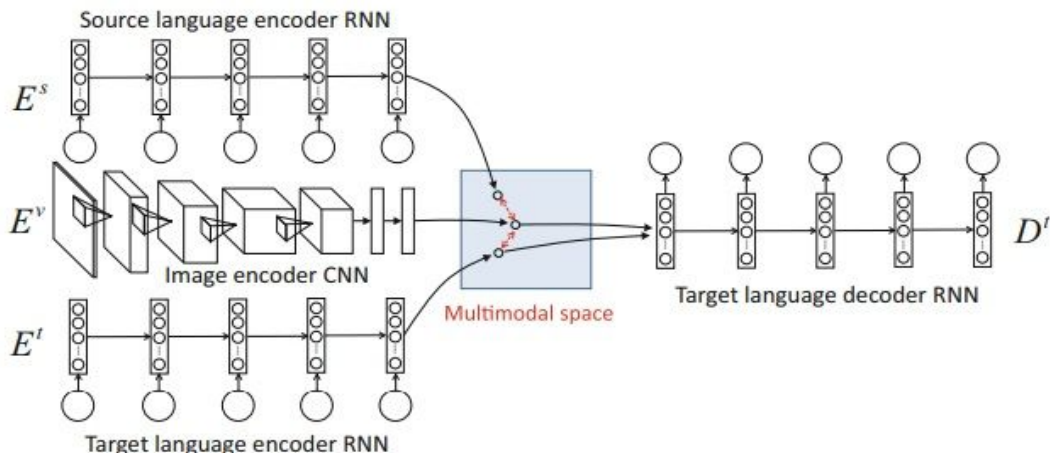


図 1：マルチモーダルエンコーダ・デコーダネットワークによる機械翻訳手法（手法 1）

項目 2：デコーダの改良による翻訳多様性の向上

手法 1 では、ソース言語とターゲット言語の完全なアラインメントをとることは原理的に不可能であるため、入力文から得られる潜在表現は必ずしもベストな翻訳に直結する情報を含んでいない。このため、潜在表現からのデコード時には、単純にトップ 1 の精度のみを求めるのではなく多様な候補を出力することが、最終的にユーザーにとって有用な翻訳結果を漏らさないために重要であると考えられる。

これを実現するために、まず、潜在表現のデコード時においてツリー構造を用いて各ステップにおける候補を管理することにより、効率的に解空間を探索するヒューリスティックアルゴリズムを提案した。本手法は従来一般的な方法であるビームサーチよりもより多様な解の候補を管理することができる（図 2）。以下、これを手法 2 - 1 と呼ぶ。

さらに、文を **semantics** や **syntax** などさまざまな観点からコード化（量子化）し、これをターゲット文の先頭に付加することで条件付きのデコーディングを行い、多様な文生成を行うアプローチを提案した。コードは出力文の意味構造や文法構造に関する抽象的なパターンに相当すると解釈でき、これを指定することで明示的に性質の異なる解候補を出力できる（図 3）。以下、これを手法 2 - 2 と呼ぶ。

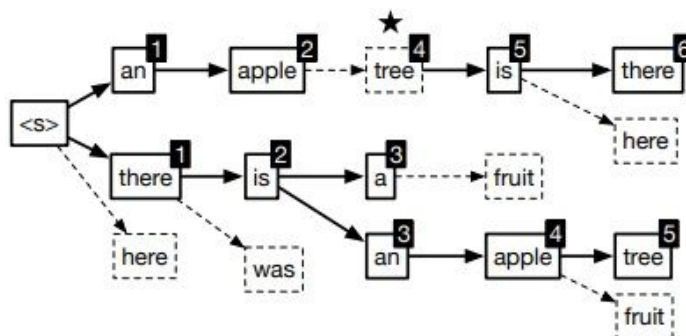


図 2：ツリー構造を用いた翻訳候補探索モデル（手法 2 - 1）

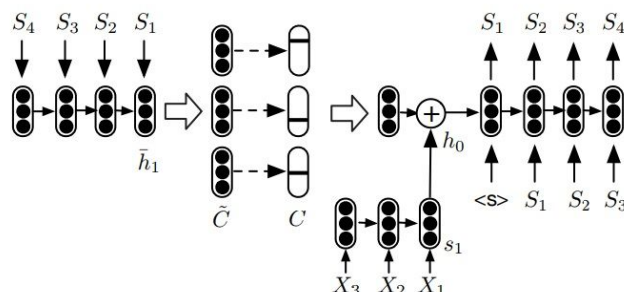


図 3：翻訳結果多様化のための離散コード学習（手法 2 - 2）

項目 3：ネットワークの高速化と軽量化

手法 1 においては、通常の機械翻訳と比較してより大量のデータを訓練に用いなければ精度の向上が見込めないため、手法の全体的な高速化と軽量化が必要となる。

まず、デコーディング時に重要となるアテンション機構における適切な窓幅を入力に応じて決定する **flexible attention** を提案し、これを導入することで計算量の削減を実現する機械翻訳手法を開発した（図 4）。以下、これを手法 3 - 1 と呼ぶ。

さらに、提案手法を大規模化するにあたってボトルネックの一つとなる単語ベクトルの容量削減を行うために、ニューラルネットワークを用いた量子化による合成的コーディング法を提案した（図 5）。より具体的には、**Gumbel-softmax** 法を応用し離散的なコードの学習を連続値の問題へ緩和することにより、コードと辞書（基底ベクトル）を両方同時に最適化するものである。以下、これを手法 3 - 2 と呼ぶ。

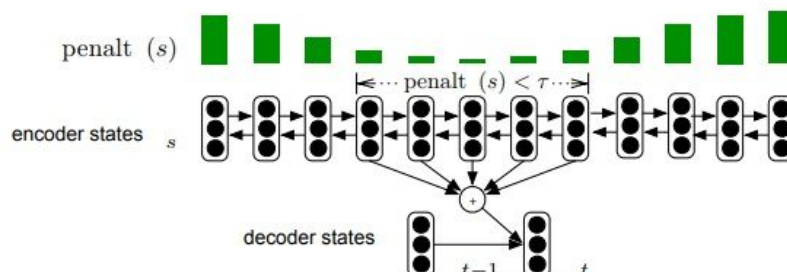


図 4：Flexible attention モデルによるデコード時の計算量削減（手法 3 - 1）

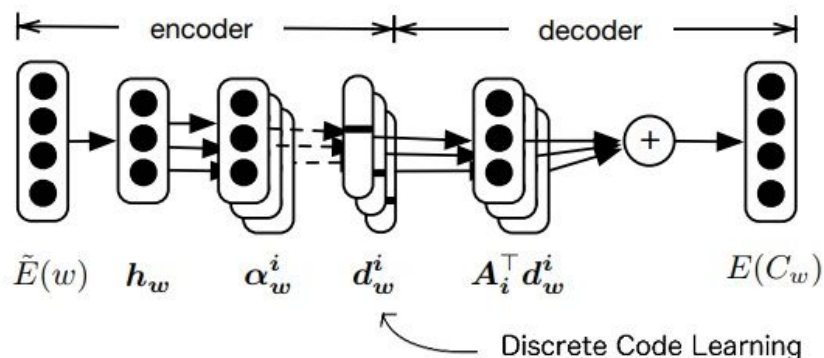


図5：構成的コーディングによる単語ベクトルの圧縮（手法3 - 2）

4. 研究成果

項目1：マルチモーダルエンコーダ・デコーダネットワークによるゼロショット翻訳の実現

手法1を実装し、英独対訳キャプションが付与された画像データセットである **IAPR-TC12** と **Multi30K** を用い、評価と定性的な分析を行った。提案手法は実際にゼロショットの翻訳が実行可能であることが示され、エンコーダの入力にターゲット言語も加えることや、学習手順の工夫により大きく性能が向上できることが判明した。一方、教師あり（対訳コーパスベース）の場合と同等の性能を得るためにはおおよそ5倍のデータを要することが示され、手法のスケラビリティを得るための高速化や軽量化が必須であることが知見として得られた。本成果は、機械翻訳の専門国際論文誌である **Machine Translation** において発表された。

項目2：デコーダの改良による翻訳多様性の向上

手法2 - 1については、**ASPEC** 英日翻訳データセットを用い、標準的な方法である **beam search** と翻訳精度・速度の両面で比較した。本手法は翻訳の速度に大きな影響を与えないが、**beam search** と比較してより多様な構造を探索できることが示され、結果として翻訳の精度も **BLEU** 値で1ポイント以上向上した。本手法は、自然言語処理のトップ国際会議である **Annual Meeting of the Association for Computational Linguistics (ACL2018)** へ採択された。

手法2 - 2については、**ASPEC** 英日翻訳データセットおよび **WMT14** 独英翻訳データセットを用い、**BLEU** 値に加え独自の多様性評価指標を定義し、類似研究との比較評価を行った。結果、本手法は翻訳の精度を維持しつつ、多様性を大きく向上できることが示された。特に、**syntax** に基づくコードを用いて翻訳を条件づけた場合に顕著な効果が得られており、直感に合致する結果となった。本手法に関する研究は、言語処理学会第25回年次大会(**NLP2019**)において最優秀賞を受賞し、**ACL2019** へ採択されている。

項目3：ネットワークの高速化と軽量化

手法3 - 1については、**ASPEC** 日英翻訳データセットと **WMT15** 独英翻訳データセットを用い、単語ベース翻訳、文字ベース翻訳のそれぞれにおいて、標準的な **attention** 手法との比較を行った。結果、単語ベース翻訳においては約60%、文字ベース翻訳においては約50%程度 **attention** の窓幅を減らせることが示された。本手法は、ニューラル機械翻訳の専門国際ワークショップである **ACL Neural Machine Translation Workshop** に採択され、**Best paper runner-up** を受賞した。

手法3 - 2については、**IWSLT 2014** 独英翻訳データセットおよび **ASPEC** 英日翻訳データセットにおいて評価し、最終的な翻訳精度を損なわずに、単語ベクトルの容量を大幅に削減できることを確認した。特に、**ASPEC** データセットにおいては99%の圧縮を実現するなど顕著な効果を得ている。また、評判分析タスクにおいても98%の圧縮を達成しており、本手法は機械翻訳に限らず自然言語処理タスク全般において使える汎用性の高い技術であることが示された。本手法に関する研究は、言語処理学会第24回年次大会(**NLP2018**)において最優秀賞を受賞し、深層学習や機械学習に関するトップ国際会議である **International Conference on Learning Representation (ICLR2018)** へ採択された。

5. 主な発表論文等

〔雑誌論文〕(計12件)

Raphael Shu, Hideki Nakayama, Kyunghyun Cho, Generating Diverse Translations with Sentence Codes, In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019 (to appear).

Raphael Shu, Hideki Nakayama, Improving Beam Search by Removing Monotonic Constraint

for Neural Machine Translation, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), pp.339-344, 2018.

Noriki Nishida, Hideki Nakayama, Coherence Modeling Improves Implicit Discourse Relation Recognition, In Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL), pp.344-349, 2018.

Raphael Shu, Hideki Nakayama, Compressing Word Embeddings via Deep Compositional Code Learning, In Proceedings of the 6th International Conference on Learning Representations (ICLR), 2018.

Noriki Nishida, Hideki Nakayama, Word Ordering as Unsupervised Learning Towards Syntactically Plausible Word Representations, In Proceedings of the Eighth International Joint Conference on Natural Language Processing (IJCNLP), pp.70-79, 2017.

Raphael Shu, Hideki Nakayama, An Empirical Study of Adequate Vision Span for Attention-Based Neural Machine Translation, In Proceedings of the First Workshop on Neural Machine Translation, pp.1-10, 2017.

Hideki Nakayama, Noriki Nishida, Zero-resource Machine Translation by Multimodal Encoder-Decoder Network with Multimedia Pivot, Machine Translation, Vol.31, Issue 1-2, pp.49-64, 2017.

〔学会発表〕(計21件)

朱中元, 中山英樹, Generating Syntactically Diverse Translations with Syntactic Codes, 言語処理学会第25回年次大会, 2019.

西田典起, 中山英樹, 談話構成素とその文脈による教師なし談話構成素構造解析, 言語処理学会第25回年次大会, 2019.

中山英樹, マルチモーダル深層学習の発展, 第20回インタラクティブ情報アクセスと可視化マイニング研究会(SIG-AM)(招待講演), 2018.

西田典起, 中山英樹, テキストの局所一貫性に基づく半教師あり暗黙的談話関係認識, 言語処理学会第24回年次大会, 2018.

朱中元, 中山英樹, 深層コード学習による単語分散表現の圧縮, 言語処理学会第24回年次大会, 2018.

西田典起, 中山英樹, Learning Syntactically Plausible Word Representations by Solving Word Ordering, 人工知能学会全国大会, 2017.

朱中元, 中山英樹, 文脈を考慮したアテンションメカニズムの計算量の削減, 人工知能学会全国大会, 2017.

Hideki Nakayama, Pivot-based Multimodality Integration for Unsupervised Cross-domain Machine Intelligence, International Symposium on Research and Education of Computational Science (RECS)(招待講演), 2017.

Hideki Nakayama, Pivot-based multimodality integration for cross-media machine intelligence, CEMS Topical Meeting on Soft Robotics(招待講演), 2016.

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等

<http://www.nlab.ci.i.u-tokyo.ac.jp/index.html>

6. 研究組織

(1)研究分担者

なし

(2)研究協力者

なし

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、研究者個人に帰属されます。