



Title	Approaches for Building an Effective System of Multi-domain Neural Machine Translation
Author(s)	羅, 文濤
Citation	大阪大学, 2020, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/77468
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏名(羅文涛ラブントウ)	
論文題名	Approaches for Building an Effective System of Multi-domain Neural Machine Translation (マルチドメインのニューラル翻訳システムの構築と評価)
論文内容の要旨	
<p>ニューラルネットに基づいた機械翻訳(ニューラル機械翻訳)は近年急速に発展を遂げた機械翻訳の技術である。既存の翻訳技術(ルールベース機械翻訳・統計ベース機械翻訳など)と比べ、ニューラル機械翻訳の訳文は翻訳精度・流暢性においてより高いレベルを達成している。ただし、このような高い翻訳品質に達する前提として、既存の翻訳技術より大量の訓練コーパスを利用して翻訳モデルを訓練することが必要である。またこの際には、翻訳対象に適した訓練コーパスを用いることが重要となる。</p> <p>翻訳対象の内容・体裁に関する分類はドメインと呼ばれる。ドメインの違いによって、訓練コーパスを構築する難度も異なる。特定のドメインにおいては訓練コーパスが量的に十分でないケースも存在し、これはニューラル機械翻訳の訳文品質を低下させる原因となる。その対策として、特定のドメインに特化した翻訳を実現するドメイン対応翻訳という翻訳方式が提案された。</p> <p>しかし、実際の翻訳への適用においては、入力原文がどのドメインに属するかさえ判明していない場合も多い。このような状況では、適用すべきドメインに適合した翻訳を実現することは難しく、従来のドメイン対応翻訳より一步前進した翻訳方式が必要とされている。</p> <p>このような課題に対し、本研究はマルチドメイン翻訳に着目したニューラル機械翻訳システムの構成法を提案する。マルチドメイン翻訳では、あらかじめ複数のドメイン対応翻訳モデルを構築しておき、翻訳対象文に応じて適切なモデルを選択・利用することにより、入力文のドメインに関わらず、ドメイン対応翻訳に等しいレベルの訳文の提供を目的とする。</p> <p>本論文では、提案方式に基づく翻訳システムの具体的な構成・実現法について述べる。また、日本語・英語・中国語を対象とした翻訳タスクによる評価結果を示し、提案方式・実現システムの有効性を検証する。本論文の各章の内容は以下のとおりである。</p> <p>第1章では、現在のニューラル機械翻訳に至る機械翻訳技術の歴史背景を概観したのち、ニューラル機械翻訳における基本技術、すなわち、エンコーダー・デコーダーモデル、アテンション機構、最新のTransformerモデルについて詳しく述べた。これらの背景を踏まえて、とくに本提案に関連するマルチドメイン翻訳方式を詳しく説明した。</p> <p>第2章では、本論が提案するマルチドメイン翻訳に着目したニューラル機械翻訳システムの全体的な構成を説明した。本システムの特徴は、あらかじめ準備する多数の翻訳モデルから翻訳対象文に応じて適合する複数のモデルを選択し、これらのモデルによる結果をアンサンブルして訳文を生成する仕組みにある。</p> <p>第3章では特定の翻訳対象ドメインに適合した翻訳モデルを訓練するために用いるドメイン適応の手法について議論した。ドメイン適応とはout-of-domainとin-domainの二種類に分けたコーパスを有効に使う考え方である。in-domainコーパスはドメインに該当するコーパスを指し、その反対に、out-of-domainコーパスはドメインに該当しないコーパスを指す。</p> <p>本章では、関連研究におけるドメイン適応の方法をデータ中心の方法とモデル中心の方法に分類して概観したのち、本研究で実施したドメイン適応の方法について述べた。具体的には、一つのドメインに対して、以下に示す5つのドメイン適応の手法を適用し、それぞれに対応する翻訳モデルを構築した。</p> <p>これらは、out-of-domainコーパスとin-domainコーパスを混合して使う「all」手法、異なる訓練の段階で異なるコーパスを使う「fine tuning」手法・「mixed fine tuning」手法・「stacking」手法、out-of-domainコーパスの一</p>	

部を選択してin-domainとして使う「data selection」手法の5つである。

第4章では本研究の中心的課題であるマルチドメイン翻訳について述べた。マルチドメイン翻訳においては、それぞれが対応するドメインに適合した複数のドメイン翻訳モデルをあらかじめ準備しておくことが前提となる。また翻訳実行時には翻訳対象文に応じて適合するドメイン翻訳モデルを複数個選択し、翻訳文の生成段階では、これらを適切のアンサンブルすることにより高品質な翻訳を目指す。

本章ではまず、構成要素となるドメイン翻訳モデルの妥当性を確認した。ドメイン適応の手法を用いない翻訳モデルを別途準備し、提案したシステムのドメイン翻訳モデルと比較した結果、提案するドメイン翻訳モデルが的確に該当ドメインに特化できたことを確認した。

次に、モデル選択とモデルアンサンブルの具体的な内容について述べた。まず、モデル選択の方法において重要なモデルの選択基準を検討した。モデルを正しく選ぶために、本論では統計指標Perplexityに基づいた選択基準を提案する。Perplexityはモデルが正解な訳文を出力する困難さを示す指標である。あるモデルが入力文に対して高いPerplexityを計量すれば、そのモデルによる翻訳の難度が高いと考えられる。つまり、そのモデルは入力文にとって不適切なモデルである。その逆の場合は、適切なモデルだと考えられる。このようにモデルの翻訳難度を測ることで、入力文のドメインが不明であっても、その文に一番適切なモデルを特定することが可能となった。

モデルアンサンブルは、上記により選択された複数の翻訳モデルを統合的に利用することにより、単独のモデルを利用する場合より質の高い訳文を生成する。この際は、各のモデルの重要度を適切に評価し、これに応じた重みを調整することにより、アンサンブルの効果を高めることが重要となる。本論ではモデルの重み分配に関して、どのモデルも平等に扱う分配法、Perplexityに基づいたランキングで重みを決める分配法の2つの手法を提案した。

第5章では各翻訳モデルに対して適用したチューニング方法について述べた。

翻訳品質を向上させるためには、ドメイン以外の側面も考慮する必要がある。本研究では、品質に影響する4つの側面を明確化し、それぞれに対して有効なチューニング方法を明らかにした。

第一の側面は「訓練コーパスの品質」である。これまでの研究で機械翻訳文を訓練コーパスとして利用することの有効性が確認されたが、機械翻訳文は多様さに欠けることで、学習しやすいデータである。

そこで、機械翻訳文にノイズを付加することで文の多様さを補強し、モデル訓練時の学習難度を上げ、訓練の質を向上させた。

第二の側面は「コーパスの構成」である。ドメイン翻訳という要素について、out-of-domainとin-domainコーパスの量を調整することでドメイン適応の効果を最適化した。

第三の問題は「単語の誤訳」である。誤訳となりやすい単語に対して置換機能を付加することで誤訳の軽減に貢献する方法を提案した。

第四の問題は「未訳語」である。未訳語は原文中の単語がそのまま訳文に出力されたものであり、ニューラル翻訳モデルのボキャブラリーに起因する問題である。通常の形態素解析により得られる単語・形態素の単位より粒度の小さいサブワードを用いることで、翻訳における言語単位の総数を減少させ、ニューラル翻訳モデルのボキャブラリーにすべての言語単位を収納することにより、未訳語を減少させることができた。

第6章では翻訳モデルの構築において重要な訓練コーパスの構築・整備について述べた。

訓練コーパスの構築においては、まず、コーパスに収録するデータの収集が必要となる。本研究では、利用可能な既存コーパス以外にウェブ上のコンテンツを収集し、ウェブスクレイピングにより、コーパスとして利用可能な形に整備した。

また、翻訳モデルの訓練に利用するために、まず文レベルの対応付けを行い、文レベルの対訳コーパスとして整備した。このために、異言語文間の類似度を計量する提案手法を適用した。

さらに、目的とする翻訳モデルにおいて悪影響を与えるかねないという対訳データの特性を分析し、これらの対訳文を排除する処理を行った。このために、翻訳対象言語における特徴を利用し、さらに既存の機械翻訳の適用を行った。

本章の最後では、構築した訓練コーパスに関する統計データを集計した。また、次章の第7章の評価タスクに即したコーパスの分類を行った。

第7章では提案手法に基づくマルチドメイン翻訳システムの有効性を確認するために行った評価実験、および、その分析結果について述べた。

具体的には日英、日中の翻訳システムを実装し、各翻訳方式の各テストセットにおける翻訳結果を考察した。評価指標としては、既存研究でも用いられているBLEUを利用した。

各翻訳方式は「候補モデルの範囲」、「モデルの最大選択数」をハイパーパラメータとして持つ。デフォルトの翻訳方式「multi_n」はマルチドメイン翻訳方式を意味し、全モデルをモデルの選択候補とする。事前に文のドメイン情報が分かる場合は、そのドメインに対応するモデルだけを選択の候補にすることが可能となる。この翻訳方式を「domain_n」と呼ぶ。ここで「n」は利用するモデルの数を指す。「multi_n」、「domain_n」が選択できるモデルの最大選択数は1から5までと設定した。テストセットに関しては、ニュース（News）、旅行（Travel）、情報技術（IT）の三つのドメインに属するセット以外に、これらの三つのドメインを混合したセット、これらのドメインとは別のドメインに属したセットも用意し、合計5つのテストセットにより評価実験を行った。

自動評価・人手評価の二種類の評価方法で訳文を評価した結果、提案手法を採用しない「baseline」翻訳方式と比べ、「multi_n」、「domain_n」のほうが質が高いことがわかった。よって、提案手法で作った機械翻訳システムの有効性が確認できた。また、「multi_n」、「domain_n」の結果が同程度であったことから、本論が提案したマルチドメイン翻訳、すなち、入力文がどのドメインに属するかわからない場合、候補の翻訳モデルから複数のモデルを選択してモデルアンサンブルを行うことで、ドメイン翻訳に等しいレベルの訳文を生成する方法が有効であることも確認できた。

一方、提案した各ドメイン適応手法について分析した結果、各ドメインにおける最適なドメイン適応法が異なるため、モデルアンサンブルを行うことが有利であることが示唆された。また、モデルの最大選択数として「domain_n」の場合、実験の設定における最大数n=5で実行するより、n=3で実行したほうが有利であった。これは、対象の翻訳タスクによって適切なモデル選択数が存在することを示唆する。

本章の最後では、多数のケース分析を行った。これらの事例から、実際の訳文において未訳語の減少、訳語選択の改善などの訳質の向上が達成されたことが定性的にも確認できた。

第8章では、本研究で得られた成果を総括し、今後の課題について整理した。

本論は、適合するドメインが事前に判明していない入力文に対しても品質の高い翻訳を実現するために、マルチドメイン翻訳システムの構成法を提案し、その有効性を評価実験により検証した。

本システムの特徴は、(1) ドメイン適応でドメインに特化した候補の翻訳モデルを複数作成しておき、(2) その中から複数のモデルを選択し、(3) 選択されたモデルのアンサンブルにより訳文を生成することにある。

翻訳品質に関する評価・分析結果から、実現したシステムの総合的な有効性だけでなく、ドメイン適応、モデル選択、モデルアンサンブルなどの各手法の有効性を確認できた。

今後の課題としては、翻訳速度の高速化、モデル選択精度の向上が重要であることを示し、これらの課題に対する研究の方向性について考察した。

論文審査の結果の要旨及び担当者

氏名 (羅文涛 (ラブントウ))		
	(職)	氏名
論文審査担当者	主査 教授	田畠智司
	副査 教授	宮本陽一
	副査 准教授	ホドシチェクボル
	副査 早稲田大学 教授	林良彦
	副査 名誉教授	岩根久

本博士学位請求論文は、「マルチドメイン翻訳に適したニューラル機械翻訳システムの構成法を提案」とともに、「同システムの有効性を検証」した論文である。ここでいう、「ドメイン」とは内容及び形式によって分類される翻訳対象の文章カテゴリーである。ニューラル機械翻訳システムの性能は、訓練において利用するコーパスに依存するため、翻訳対象のドメインと訓練コーパスのドメインが不整合である場合、適切な訳出ができない場合がある。この問題を解決するため、特定のドメインに特化したドメイン対応翻訳方式が提案された。しかし、この方式においては通常は翻訳対象となる原文のドメインを特定する過程が必要になるため、実際の使用において適用可能な場面が制限されてしまうという問題があった。そこで本論文では、入力文のドメインを問わず、ドメイン対応翻訳に等しいレベルの翻訳を行えるマルチドメイン翻訳方式を提案し、ニューラル機械翻訳技術に基づくシステムの構成法を具体的に示すとともに、その有効性を実験的に検証している。

本論文は全8章からなる。まず、第1章では、従来のルールベース型機械翻訳・統計的機械翻訳など、本研究の背景であるニューラル翻訳技術に至る機械翻訳技術の発展について記述したのち、ニューラル機械翻訳技術において基本となる技術、すなわち、エンコーダー・デコーダーモデル、アテンション機構、Transformerモデルについて詳述している。これら技術の変遷、進化をふまえ、本研究が提案するマルチドメイン翻訳方式を導入し、その位置付けを明確に述べている。

第2章では、本論文が提案するマルチドメイン・ニューラル機械翻訳システムの構成について論じている。本システムの大きな特徴となるメカニズムは、あらかじめ用意した多数の翻訳モデルから翻訳対象文に適合する複数のモデルを選択し、これらの複数のモデルによる翻訳結果を適切に組み合わせるアンサンブルを行なって訳文を生成することである。

第3章では特定のドメインに適合した翻訳モデルを訓練するために用いるドメイン適応の手法について議論している。ドメイン適応には、訓練用コーパスを、翻訳対象とは異なるドメインに属する「out-of-domain コーパス」と翻訳対象と合致する「in-domain コーパス」に分類することによってモデルの精緻化を行う方法があることを解説する一方、「データ中心のモデル適応方法」と「モデル中心の適応方法」に分類し知見の整理・集約化を試みている。本提案では、具体的には1つのドメインに対して、以下に示す5つのドメイン適応の手法を適用し、それぞれに対応する翻訳モデルを構築して比較検討している。それらは、out-of-domain コーパスと in-domain コーパスを混合して使う「all」法、異なる訓練の段階で異なるコーパスを使う「fine tuning」法・「mixed fine tuning」法・「stacking」法、out-of-domain コーパスの一部を選択して in-domain として使う「data selection」法の5つである。

第4章では中心的課題であるマルチドメイン翻訳の技術的側面について述べている。まず、第3章で提案したドメイン適応の手法によって構築した複数のドメイン翻訳モデルが的確に該当ドメインに特化できることを実験により検証している。次に、統計指標 Perplexityに基づく評価基準を適用することにより翻訳入力文に適合するモデル選択を行うことの妥当性について詳細な検討を行なっている。翻訳モデルの出力文の Perplexity が高いことは、その文が目的言語の文としての尤もらしさが低いことを示す。このため、所与のモデルが入力文に対して正解とみなしうる訳文を出力する困難度を示す指標として Perplexity を利用することができる。この統計指標をもとに、選択された複数の翻訳モデルを統合的に組み合わせ、単独モデルと比較してより質の高い訳文を生成するモデルアンサンブル手法が本論文の核となる技術である。さらに、アンサンブルを行う際、各モデルの重み付け分配による結果の差異に着目し、どのモデルも均等に扱う分配法、Perplexity にもとづいたランキングで重みを決める分配法の2つの手法を比較検討している。

第5章は、翻訳品質の向上を図るために、以下にあげる4つの視点から各翻訳モデルに対するチューニング方法を実装した結果を述べている。

- 「訓練コーパスの品質」：訓練用コーパスにノイズを付加することでモデル訓練時の学習難度を上げ、訓練

の質を向上させる調整法。

2. 「コーパスの構成」：out-of-domain と in-domain コーパスの分量比を調整することでドメイン適応効果を最適化する調整法。
3. 「単語の誤訳」：誤訳となりやすい単語に対して置換機能を付加することで誤訳の低減を図る調整法。
4. 「未訳語の軽減」：原文中の単語がそのまま訳文に出力される未訳語の問題は、ニューラル翻訳モデルのボキャブラリーがすべての言語単位を収納できていないことに由来するものである。通常の単語・形態素の単位より粒度の細かいサブワードを用いることで、翻訳における言語単位の総数を減少させ、未訳語を減少させる調整法。

第6章は、翻訳モデルの構築において重要な訓練コーパスの構築・整備について述べたものである。本研究では、まず、コーパスのデータ収集について説明している。利用可能な既存コーパス以外に、ウェブ上のコンテンツを収集し、利用可能な形で整備している。この際は、翻訳モデルの訓練に利用するために、データを文レベルの対訳コーパスとして整備した上で、文レベルの対応付けを異言語文間の類似度に基づいて行っている。さらに、翻訳モデルに悪影響を与える可能性のある対訳データを分析し、排除する処理を行っている。本章の最後では、これらの処理により構築した訓練コーパスの分類・集計結果を、次章で述べる評価タスクに即した形で述べている。

第7章では日本語・英語・中国語を対象とした翻訳タスクを設定し、提案手法に基づくマルチドメイン翻訳システムの有効性を評価実験により確認している。また、翻訳事例の詳細な評価分析を実施している。自動評価と人手による評価を比較考察した結果、提案手法を採用しない「baseline」翻訳方式に比べ、提案手法を利用した翻訳方式「multi_n」、「domain_n」（nは選択・利用する翻訳モデルの数）の方が質が高いことを明らかにし、本論文が提案する機械翻訳システムの有効性を確認している。他方、「multi_n」、「domain_n」の結果はほぼ同程度であったことを報告している。デフォルトのマルチドメイン翻訳方式「multi_n」は全モデルをモデルの選択候補とするのに対し、翻訳方式「domain_n」は、事前に文のドメイン情報を明示し、ドメインに対応するモデルだけを選択候補とする。よって、入力文がどのドメインに属するかわからない場合に、本論文が提案する機械翻訳システムが有効であることを確認している。一方、提案した各ドメイン適応手法の評価分析を行った結果、News, Travel, IT の3つのドメインに対する最適化の方法が異なるため、複数モデルのアンサンブルを行うことが有効であるということを示している。さらに、翻訳方式「domain_n」の場合、モデル選択数 $n=5$ の実行に比して、 $n=3$ で実行した方が有利な結果が得られたと報告している。これは、翻訳タスクによって適切なモデル選択数が異なることを示唆している。これらの定量的な評価に加え、本章の最後では多数のケース分析を行い、実際の訳文において未訳語の減少、訳語選択の改善などの点で訳質の向上が達成されたことを定性的に確認するとともに、今後の研究の方向性の示唆を与える課題点を抽出している。

最終第8章は、本研究で得られた成果を総括し、今後の課題について言及した章となっている。本論文が学界、産業界に対して成し得る新たな貢献は、適合するドメインが事前に判明していない入力文に対して高品質の翻訳を実現するために、マルチドメインのニューラル翻訳システムの構成法を提案したこと、および、その有効性を評価実験により検証したことである。本論文で詳述したマルチドメイン・ニューラル翻訳システムの特徴は下記の3点に集約される。

- (1) ドメインに適切に特化した翻訳モデルを構築するためのドメイン適応手法の確立。
- (2) 翻訳対象文に適合する複数の翻訳モデルを選択するための統計的指標の利用法の確立。
- (3) 選択された複数の翻訳モデルのアンサンブルにより精度の高い訳文を生成できるシステムの実現。

翻訳品質に関する評価・分析結果から、実現したシステムの総合的な有効性だけでなく、ドメイン適応、モデル選択、モデルアンサンブルなどの各手法の有効性を確認した上で、今後の課題として、翻訳速度の高速化、モデル選択精度の向上が重要であることを提示し、これらの課題に対する研究の方向性を示し論文を締めくくっている。

本論文には、翻訳モデルのさらなる精緻化の余地や、実際の使用事例での翻訳文の質の定量的、定性的測定法の吟味など、いくつか未実行の課題もあるが、これらは本論文の価値を損なうものではなく、今後の研究によるさらなる発展が期待されるものである。全体として本論文は、複数のドメインに同時適応可能なニューラル機械翻訳システムの構成法を論じた点で機械翻訳システムの実用性の向上に寄与する優れた研究であるとともに、先端の自然言語処理技術と言語文化学を融合させた意義深い研究であり、その学術的貢献は明確である。

以上のように、本論文を博士（言語文化学）の学位論文として価値のあるものと認める。
なお、チェックツール iThenticate を使用し、剽窃、引用漏れ、二重投稿等のチェックを終えていることを申し添えます。