

原言語側の欠落を考慮した Multi-Source NMT

西村 優汰 須藤 克仁 中村 哲

奈良先端科学技術大学院大学 情報科学研究科

{nishimura.yuta.nn9, sudoh, s-nakamura}@is.naist.jp

1 はじめに

機械翻訳では学習に大規模な対訳コーパスを必要とし、翻訳精度は対訳コーパスのデータ量に大きな影響を受けるため、十分な資源がない言語対では高精度な機械翻訳の実現が難しい。このような問題を解決する手法として、大規模コーパスがある言語対と組み合わせて翻訳を行う多言語機械翻訳 [1][2][3] が存在する。これらの研究では、原言語や目的言語の言語数に応じて、エンコーダやデコーダの数を調整したり、エンコーダとデコーダを1つずつしか用いないが全てのパラメータを全言語で共有したりすることで多言語機械翻訳を実現している。多言語機械翻訳では、単言語対での機械翻訳よりも高い精度で機械翻訳を行えることが示されている。

一方で、多言語機械翻訳を行うコーパスでは、ある文に対して全ての言語に対応する文が揃っている状況は非常に限定される。このような複数言語が存在するコーパスにおいて対訳文が完全に揃っていないものを本稿では欠落コーパスと呼び、また、コーパス中の全ての言語において対訳文が揃っているものを完全コーパスと呼ぶ。特に従来の原言語が複数、目的言語が1つの多言語機械翻訳 (Multi-Source NMT)[1] では、完全コーパスからの学習を前提としており、コーパス中の欠落部分における対訳文は学習に活かされていない。しかし、欠落コーパスにおいても、対訳が存在する一部の言語対の情報から翻訳の学習は可能であると考えられる。そこで本研究では、欠落コーパスにおいて対訳が全て揃っていない部分も学習に用いる Multi-Source NMT について提案する。極めて単純な手法でより精度の高い Multi-Source NMT が行えることを実験により確認した。

2 Many-to-One NMT

複数の原言語を用いる NMT のアプローチとして、複数の原言語を翻訳器に入力する Multi-Source NMT と、異なる原言語ごとに作成した目的言語への翻訳器

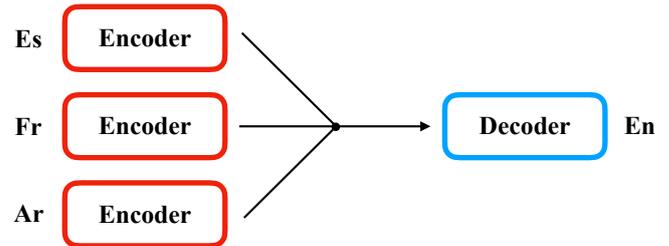


図 1: Multi-Source NMT

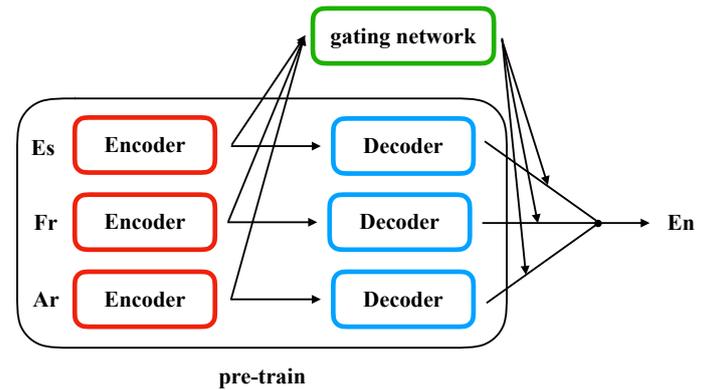


図 2: Mixture of NMT Experts

の翻訳結果を統合する Mixture of NMT Experts の 2 つが提案されている。本節ではそれぞれのアプローチについて紹介する。

2.1 Multi-Source NMT

Zoph ら [1] は、原言語を複数、目的言語を 1 つ (Many-to-One) 用いた Multi-Source NMT を提案している。この手法では、図 1 で示すようにエンコーダを原言語の数、デコーダを 1 つ用いることで機械翻訳を行なっている。Multi-Source NMT を行うことで単言語対での機械翻訳よりも精度向上に有効であることが示されている。しかし、Multi-Source NMT では全言語対の対訳が揃っていることが前提であり、ある 1 つの言語の対訳文が欠落しているとその文は学習に用いることができない。

ここで、Multi-Source NMT におけるエンコーダとデコーダの接続部分について簡潔に述べる。原言語が2種類ある場合を考える。エンコーダは2つあるので、エンコーダの最終 hidden 層をそれぞれ h_1, h_2 とし、cell も同様に c_1, c_2 とする。この時、デコーダの hidden 層の初期状態 h 、cell の初期状態 c は次のように与えられる。

$$h = \tanh(W_c[h_1; h_2]) \quad (1)$$

$$c = c_1 + c_2 \quad (2)$$

また、Attention についても簡潔に述べる。タイムステップ t の時の各エンコーダの context vector を c_t^1, c_t^2 、デコーダの hidden 層を h_t とした時のデコーダの softmax 層に送る前の最後の hidden 層 \tilde{h}_t は次のように与えられる。

$$\tilde{h}_t = \tanh(W_c[h_t; c_t^1; c_t^2]) \quad (3)$$

2.2 Mixture of NMT Experts

欠落コーパスであっても、全ての対訳文を用いて学習を行う手法としてはアンサンブル学習を用いた手法が存在する。Garmash ら [4] は、原言語が複数、目的言語が1つの状況において、まず、原言語の数だけ Single モデルを作成し pre-train を行い、その後、各モデルを各エキスパートとして扱った混合エキスパートモデルにおいてアンサンブルする手法を提案している。この手法の概略を図2に示す。この手法では、それぞれの Single モデルの学習が独立しているので完全コーパスを前提とする必要がなく、欠落コーパスであっても不揃い部分を用いて学習を行うことができる。ただし、エキスパートに重みをかける gating network を学習させるには、全言語が揃っている部分が少量でも必要である。

ここで、Mixture of NMT Experts のアンサンブル部分について簡潔に述べる。原言語の数を m とし、pre-train を行なった j 番目の Single モデルのデコーダの入力を $f^j(x)$ 、softmax 層の出力を y_j とした時に、アンサンブルした出力 y^e は次のように与えられる。

$$y^e = \sum_m g_j y_j \quad (4)$$

g_j は j 番目の Single モデルの出力層における gating network から算出された重みである。 g は次の式で求められる。

$$g = \text{softmax}(W_{gate} \tanh(W_{hid}[f^1(x); \dots; f^m(x)])) \quad (5)$$

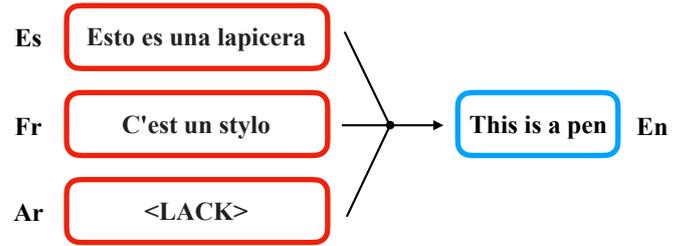


図 3: 提案法の例

3 提案法：原言語側の欠落を考慮した Multi-Source NMT

本研究では、Multi-Source NMT の精度を向上させるために、欠落コーパスであっても利用可能な対訳文は余すところなく学習に利用する手法を提案する。

コーパス中の欠落している部分に欠落部分であることを知らせる特殊記号 “<LACK>” を文の代わりに挿入するという極めて単純な手法によって欠落部分を埋めた。これは、欠落部分を無視しつつ、訳文が存在するところだけで翻訳できるように学習が進むことを期待したものである。ただし、Multi-Source NMT の学習手法については従来の手法と同じ手法を用いた。

図3を用いて提案手法について簡潔に説明する。原言語がスペイン語、フランス語、アラビア語の3言語、目的言語が英語の1言語であり、アラビア語の文が欠落している場合を想定する。この場合、アラビア語以外のエンコーダにはそれぞれの対訳文を入れ、アラビア語のエンコーダには特殊記号 “<LACK>” を入れて Multi-Source NMT を行う。

4 実験

4.1 欠落コーパス

本研究では、コーパスが欠落していることによって機械翻訳の精度がどの程度悪化するかについても調査するため、コーパスそのものが欠落コーパスであるものを用いるのではなく、完全コーパスから擬似的な欠落コーパスを作成してシミュレーション実験を行なった。欠落コーパスの作成方法としては、完全コーパスから一部の文章を削除することによって、擬似的な欠落コーパスを作成した。

4.2 実験データ

コーパスは UN6WAY [8] を用いた。Many-to-One の多言語機械翻訳を行うにあたって、原言語をスペイン語、フランス語、アラビア語の3言語とし、目的言語を英語の1言語とした。UN6WAY のコーパスから最

表 1: 欠落コーパスの設定 (×は欠落を示す)

sentence No.	Spanish	French	Arabic	English
1-200,000	×	○	○	○
200,001-400,000	○	○	×	○
400,001-600,000	○	×	○	○
600,001-800,000	○	○	○	○
800,001-1,000,000	○	○	○	×

大文長を 40 語として、上から 100 万文を抽出したものを完全コーパスとし、この完全コーパスから一部の文を削除することにより欠落コーパスを作成した。欠落コーパスの設定を表 1 に示す。表 1 中の “sentence No.” はコーパスにおける行番号を表しており、“×” 部分が完全コーパスから削除した部分である。また、開発データとして 1000 文、テストデータとして 4000 文を訓練データとは別に UN6WAY のコーパス中から抽出した。

4.3 実験内容

実験には、Multi-Source NMT, Mixture of NMT Experts, Single NMT[5] の 3 つのモデルを用いた。コーパスには、完全コーパスと欠落コーパスの 2 つのコーパスを用いた。

4.3.1 完全コーパスを用いた実験

まず、完全コーパスでの実験として、100 万文全てを用いて Multi-Source NMT, Single NMT での実験を行い、学習済みの Single NMT を用い Mixture of NMT Experts の gating network の学習を行なった。

4.3.2 欠落コーパスを用いた実験

次に、欠落コーパスでの実験について述べる。Multi-Source NMT においては、従来の手法であれば、全言語対の対訳文が揃っている 20 万文 (sentence No. 600,001-800,000) でしか学習ができないが、原言語コーパスの欠落部分に特殊記号 “<LACK>” を挿入することで 80 万文を用いて実験を行なった。Single NMT においては、各言語対それぞれ 60 万文の対訳データを用いて実験を行った。この学習済みの Single NMT を用いて、全言語対の対訳文が揃っている 20 万文を用いて Mixture of NMT Experts の gating network の学習を行なった。

4.4 実験設定

実験に用いた NMT モデルは、Luong ら [5] によって提案された Global Attention と Attentional Feeding を使用し、さらにエンコーダでは、Bahdanau らの手法

に使用された Bidirectional Encoder を使用した。エンコーダとデコーダは、2 層の LSTM を使用し、hidden 層と embed 層のユニット数はそれぞれ 512 とした。語彙数は、原言語の各言語、目的言語共に訓練データ中の頻出単語 30000 語とした。モデルの最適化のアルゴリズムとして Adam を使用し、gradient clipping を 5 に設定した。また、Mixture of NMT Experts での gating network における hidden 層のユニット数は 256 とした。

評価手法として、BLEU[6] を使用し、開発データで Log Perplexity が最小となった時点でのパラメータを保存しテストデータで評価した。

5 実験結果

表 2 に Multi-Source NMT と Mixture of NMT Experts の結果を示す。また、表 3 に Single NMT の結果を示す。

まず、表 2,3 全体をみてわかることとしては、どのモデルにおいても訓練データの文数が少なくなると BLEU が下がっていることがわかる。完全コーパスと欠落コーパスでの BLEU を比較すると、Single モデルでは、約 1 ポイント、Multi-Source NMT では約 6 ポイント、Mixture of NMT Experts では約 2 ポイント下がっている。Many-to-One でのモデルでは、Single モデルに比べて、BLEU の落ち幅は大きい。これは、完全コーパスと欠落コーパスにおける訓練データの文数の違いに起因するものだと考えられる。全てのモデルにおいて完全コーパスでは 100 万文の訓練データを用いていたが、Single モデルでは 60 万文、Multi-Source NMT, Mixture of NMT Experts では 20 万文と訓練データの文数に大きな差がある。

しかし、Single モデルで最も高い BLEU を出している Spanish → English のモデルと Many-to-One のモデルを比較すると、完全コーパス、欠落コーパスのどちらの場合でも、Many-to-One のモデルの方が Single モデルよりも BLEU が高い。欠落コーパスを用いた Multi-Source NMT の訓練データの文数は 20 万文、Single モデルでは 60 万文であることから、訓練データの文数がある程度少なくとも Multi-Source NMT の方が良い性能であることがわかる。

さらに、Multi-Source NMT において欠落コーパスの欠落部分を特殊記号 “<LACK>” で置換して訓練データを余すことなく用いる提案手法では、欠落コーパスを用いたモデルの中では BLEU が最も高いことがわかる。欠落コーパスにおいて欠落があっても可能な限りデータを用いる場合と欠落部分が存在したらその

表 2: Many-to-One での実験結果

	Multi-Source NMT			Mixture of NMT Experts	
	完全 (100 万文)	完全 (20 万文)	欠落 (80 万文:提案法)	完全 (100 万文)	欠落 (80 万文)
BLEU	37.53	31.24	36.43	34.10	32.16

表 3: Single での実験結果

	原言語	完全 (100 万文)	欠落 (60 万文)
BLEU	Spanish	32.09	30.98
	French	26.20	25.62
	Arabic	23.62	22.02

文全て無視をしてしまう場合での Multi-Source NMT の BLEU の差が約 5 ポイントもあることがわかる。このことから、欠落部に特殊記号を置換するというシンプルな提案手法は、不揃い部分を活かす手法としては極めて有効であると考えられる。

また、Mixture of NMT Experts よりも Multi-Source NMT の方が BLEU が高いことから、原言語それぞれの意味空間を作るよりも共通の意味空間を作った方がモデルの性能が上がるのではないかと考えられる。ただし、原言語の種類、数によっても変化すると考えられるので、最適な言語の種類、数も今後調査していきたい。

6 おわりに

本研究では、欠落コーパスを用いて Multi-Source NMT を行う際に、欠落部分を特殊記号で置換するという極めてシンプルな手法を用いてコーパス中の不揃い部分を活かすことで翻訳精度がより高くなることを示した。

本研究で用いた欠落コーパスは、任意の文において 4 言語中 1 言語のみが欠落している状況を想定して作成したが、他に可能な状況としては、複数言語が欠落している場合、テストデータにも欠落がある場合など様々な状況が想定される。今後、様々な状況を想定した欠落コーパスを作成し、さらにシミュレーション実験を行なっていく。その後、TED Talks の字幕翻訳から作成された対訳コーパスである WIT³[9] などの実際に欠落コーパスであるものを用いて実験を行なっていく。また、欠落コーパスを用いた実験として、Many-to-One の多言語機械翻訳だけでなく、One-to-Many, Many-to-Many でも実験を行なっていく予定である。

謝辞

本研究の一部は JSPS 科研費 JP17H06101 の助成を受けたものである。

参考文献

- [1] Zoph, B., and Knight, K. Multi-source neural translation. In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17 (2016).
- [2] Dong, D., Wu, H., He, W., Yu, D., and Wang, H. Multi-task learning for multiple language translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (2015).
- [3] Melvin, J., Mike, S., Quoc, L., Maxim, K., Yonghui, W., Zhifeng, C., Nikhil, T., Fernanda, V., Martin, W., Greg, C., Macduff, H., and Jeffrey D. Google's multilingual neural machine translation system: Enabling zero-shot translation. CoRR, abs/1611.04558 (2016).
- [4] Garmash, E., and Monz, C. Ensemble learning for multi-source neural machine translation. In Proceedings of COLING (2016).
- [5] Luong, M.-T., Pham, H., and Manning, C. D. Effective approaches to attention-based neural machine translation. In Conference on Empirical Methods in Natural Language Processing (2015).
- [6] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. BLEU: a method for automatic evaluation of machine translation. In ACL, (2002).
- [7] Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proceedings of ICLR (2015).
- [8] Ziemski, M., Juncys-Dowmunt, M., and Pouliquen, B. The united nations parallel corpus v1.0. In LREC (2016).
- [9] Cettolo, M., Girardi, C., and Federico, M. WIT3: Web inventory of transcribed and translated talks. In Proceedings of EAMT (2012).