

固有表現を考慮したニューラル機械翻訳 におけるリランキングの活用

南端 尚樹[†] 田村 晃裕[†] 加藤 恒夫[†]

[†]同志社大学大学院

1 はじめに

ニューラル機械翻訳 (NMT) では、原言語や目的言語の固有表現 (NE) の情報を活用することで翻訳性能が改善されている。これまで、原言語及び目的言語の NE を活用する NMT モデルとして、タグ付けモデル[1]や埋め込みモデル[2]、これら二つのモデルをアンサンブルするアンサンブルモデル[3]などが提案されている。

NMT においては、出力候補文をリランキングすることで翻訳性能が改善できる場合がある[4, 5]。しかし、従来の NE を活用する NMT モデルではリランキングが活用されていない。そこで本研究では、NE に基づく NMT にリランキングを導入する。具体的には、NE に基づく NMT のためのリランキングとして、原言語文中の NE に対する逆翻訳の尤度に基づくリランキングを提案する。そして、NE に基づく NMT において、従来の逆翻訳の尤度[4]や Masked Language Model (MLM) の尤度[5]に基づくリランキング及び提案のリランキングの効果を検証する。WMT2020 の英語と日本語間の翻訳実験において、NE に基づく NMT モデルであるアンサンブルモデル[3]にリランキングを導入することで翻訳性能が向上することを確認した。また、日英翻訳では、提案リランキングが従来の逆翻訳の尤度に基づくリランキングよりも翻訳性能を改善できることを確認した。

2 NE に基づく従来 NMT : アンサンブルモデル

本節では、本研究でリランキングの導入対象としたアンサンブルモデル[3]を説明する。アンサンブルモデルは、独立に学習したタグ付けモデル[1]と埋め込みモデル[2]を推論時にアンサンブルすることで翻訳を行う。具体的には、各時刻 t において、目的言語の単語と NE の対 o_t に対するタグ付けモデルの出力確率と埋め込みモデルの出力確率を平均した確率に基づいて目的言語文を生成する。 o_t に対するアンサンブルモデルの出力確率は次の通りである。

$$p(o_t | \mathbf{x}, \mathbf{o}_{1:t-1}) = \frac{p_{tag}(o_t | \mathbf{x}, \mathbf{o}_{1:t-1}) + p_{emb}(o_t | \mathbf{x}, \mathbf{o}_{1:t-1})}{2}$$

ここで、 \mathbf{x} は NE の情報が付いた原言語文、 $p_{tag}(o_t | \mathbf{x}, \mathbf{o}_{1:t-1})$ はタグ付けモデルの出力確率、 $p_{emb}(o_t | \mathbf{x}, \mathbf{o}_{1:t-1})$ は埋め込みモデルの出力確率である。

Named Entity-based Neural Machine Translation with Reranking

Naoki Minamibata[†] (ctwh0141@mail4.doshisha.ac.jp)

Akihiro Tamura[†] (aktamura@mail.doshisha.ac.jp)

Tsuneco Kato[†] (tsukato@mail.doshisha.ac.jp)

[†] Graduate School, Doshisha University

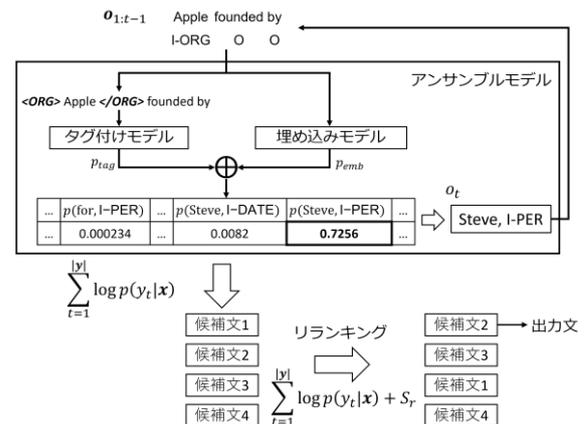


図 1: リランキングを導入したアンサンブルモデル

る。タグ付けモデルは、原言語文及び目的言語文中の NE の前後に、NE の種類と開始/終了情報を含む NE タグを挿入することで NE を考慮する。一方の埋め込みモデルは、エンコーダ及びデコーダの埋め込み層で、単語の埋め込み表現に NE の埋め込み表現を加えることで NE を考慮する。

3 提案手法: リランキングの導入

リランキングを導入したアンサンブルモデルの概要を図 1 に示す。従来の NMT モデルでは、対数尤度が最も高い目的言語文を出力する。一方、リランキングを導入したモデルでは、まず、対数尤度の高い上位 N 個の出力候補文を生成する。その後、各出力候補文に対して、目的言語文の対数尤度とは異なる観点のスコア (S_r) を加えたリランキングスコア (S) を計算し、算出したリランキングスコアが最も高い目的言語文を出力とする。リランキングスコアは次の通りである。

$$S = \sum_{t=1}^{|y|} \log p(y_t | \mathbf{x}) + S_r$$

ここで、 $|y|$ は目的言語文 \mathbf{y} の単語数、 S_r はリランキングに用いるスコアである。なお、本研究では、スコア間のスケールの違いを考慮し、各スコアは標準化してから足し合わせる。以降では、本研究で検討する S_r を説明する。

3.1 逆翻訳の尤度 (S_{BT})

S_{BT} は、翻訳対象の言語方向 ($s \rightarrow t$) とは逆方向 ($t \rightarrow s$) の逆翻訳モデルを用いて、目的言語文 \mathbf{y} から原言語文 \mathbf{x} を生成させたときの対数尤度である。 S_{BT} は次式の通り算出する。

$$S_{BT} = \sum_{t=1}^{|\mathbf{x}|} \log p_{t \rightarrow s}(x_t | \mathbf{y})$$

ここで、 $|\mathbf{x}|$ は原言語文 \mathbf{x} の単語数である。この S_{BT}

表 1：翻訳性能 (BLEU (%))

モデル	英日	日英
BS	23.03	22.70
BS+S _{BT}	23.15	22.75
BS+S _{MLM}	22.85	22.90
BS+S _{NE}	23.06	22.79
BS+S _{MLM} + S _{BT}	23.28	23.01
BS+S _{MLM} + S _{NE}	23.10	23.11

が大きいほど、目的言語文は原言語文の正しい翻訳になっていると考えられる。

3.2 MLM の尤度 (S_{MLM})

S_{MLM}は、MLM による目的言語文 \mathbf{y} の対数尤度である。具体的には、 \mathbf{y} 中の単語を一単語ずつマスクし、マスク部分を MLM で予測させることで \mathbf{y} 中の各単語の対数尤度を求める。その対数尤度の和が S_{MLM}である。S_{MLM}は次式の通り算出する。

$$S_{MLM} = \sum_{t=1}^{|\mathbf{y}|} \log p_{MLM}(y_t | \mathbf{y}_{\setminus t}; \Theta)$$

ここで、 Θ は MLM のパラメータである。この S_{MLM}が大きいほど、目的言語文は自然な文と考えられる。

3.3 原言語文中の NE に対する逆翻訳の尤度 (S_{NE})

S_{NE}は、逆翻訳モデルを用いて、目的言語文 \mathbf{y} から原言語文中の NE を生成させたときの対数尤度である。S_{BT}のうち NE のみに着目したスコアともいえる。S_{NE}は次式の通り算出する。

$$S_{NE} = \sum_{w \in C} \log p_{t \rightarrow s}(w | \mathbf{y})$$

ここで、Cは原言語文 \mathbf{x} 中の NE の集合である。この S_{NE}が大きいほど、目的言語文は原言語文中の NE を正しく翻訳できていると考えられる。

4 実験

本実験では、WMT2020 の英語と日本語間のニューズ翻訳タスク (教師データ：7,455,119 文対、開発データ：1,774 文対、評価データ：1,000 文対) において、アンサンブルモデル[3] (ベースライン：BS) に3節のランキングを導入した際の翻訳性能を調査する。アンサンブルモデルのハイパーパラメータの設定や学習方法は文献[3]に従った。翻訳性能の評価指標には BLEU (%) を用いた。ランキング時の出力候補文数 (N) は 4 とした。S_{MLM}の算出では MLM として RoBERTa を使用した。

実験結果を表 1 に示す。表 1 より、英日翻訳の S_{MLM}を除いて、S_{BT}、S_{MLM}、S_{NE}を用いてランキングすることでアンサンブルモデルの翻訳性能が向上することを確認した。また、逆翻訳に基づくスコア (S_{BT}、S_{NE}) と MLM に基づくスコア (S_{MLM}) の両方を用いてランキングすることで、翻訳性能がさらに向上することを確認した。

逆翻訳に基づくスコアである S_{BT} と S_{NE} を比較すると、英日翻訳では S_{BT} の方が性能向上に寄与し、日

表 2：出力候補文数 (N) を変えた際の翻訳性能 (日英翻訳)

モデル	N=4	N=10	N=20
BS+S _{BT}	22.75	22.61	22.74
BS+S _{MLM}	22.90	23.08	23.27
BS+S _{NE}	22.79	22.81	22.97
BS+S _{MLM} + S _{BT}	23.01	23.06	23.13
BS+S _{MLM} + S _{NE}	23.11	23.13	23.37

英翻訳では S_{NE} の方が性能向上に寄与する結果となった。最終的に、ランキングによって英日翻訳では最大 0.25 ポイント、日英翻訳では最大 0.41 ポイント BLEU が向上した。

5 考察

本節では、日英翻訳において、ランキング時の出力候補文数 (N) を変更した際の翻訳性能の変化を調査した。N=4、N=10、N=20 の場合の BLEU を表 2 に示す。表 2 より、BS+S_{BT}を除いて、N が大きくなるほど BLEU が高くなるのが分かる。これより、出力候補文の文数を大きくするほど、ランキングの効果も大きくなる傾向があることを確認した。

6 まとめ

本研究では、NE に着目したランキングを提案し、提案ランキングや従来の逆翻訳に基づくランキング、MLM に基づくランキングを NE に基づく NMT モデルに導入した。WMT2020 の英語と日本語間の翻訳実験において、ランキングを導入することでアンサンブルモデルの翻訳性能が向上することを確認した。そして、出力候補文の文数を大きくするほどランキングの効果も大きくなることを確認した。さらに、日英翻訳では、提案ランキングが従来の逆翻訳の尤度に基づくランキングよりも翻訳性能を改善できることを確認した。

参考文献

- [1] 南端 尚樹, 田村 晃裕, 加藤 恒夫. 目的言語文の固有表現タグ付与に基づく Transformer ニューラル機械翻訳. NLP2022 発表論文集, pp.937-941, 2022.
- [2] A. Siekmeier, W. Lee, H. Kwon, and J. -H. Lee. Tag Assisted Neural Machine Translation of Film Subtitles. In Proc. of IWSLT, pp.255-262, 2021.
- [3] 南端 尚樹, 田村 晃裕, 加藤 恒夫. タグ付けモデルと埋め込みモデルのアンサンブルに基づき固有表現を考慮するニューラル機械翻訳モデル. 情報処理学会 NL 研, 2023-NL-257 (2), pp.1-9, 2023.
- [4] 後藤 功雄, 田中英輝. ニューラル機械翻訳での訳抜けした内容の検出. 自然言語処理, 25(5), pp.577-597, 2018.
- [5] J. Salazar, D. Liang, T. Q. Nguyen, and K. Kirchoff. Masked Language Model Scoring. In Proc. of ACL, pp.2699-2712, 2020.