多数決による自己回帰モデルに基づく機械翻訳

村上仁一 [1]
[1] 鳥取大学工学部 murakami@tottori-u.ac.jp

概要

ニューラルネットワーク機械翻訳（以後 NMT）は，現在の機械翻訳において主流である．NMT は，encoder-decoder モデルに基づいている．この encoder-decoder モデルは，自己回帰モデル（以後 AR モデル）に極めて類似している．

そこで，NMT において AR モデルを仮定し，この仮定の下で，機械翻訳システムを構築した．具体的には，従来の NMT において，AR モデルを想定して，多数決で翻訳文を選択した．また，入力文ごとに少数の類似文を用いた再学習をおこなった．その結果，少量の学習データにも関わらず，現在の google に接近した，高い翻訳性能が得られた．

1 はじめに

ニューラルネットワーク翻訳（以後 NMT）は，基本的には，encoder-decoder モデル [1] を利用している．基本的な考え方は，入力文を $j_1j_2j_3$ として，出力文を $e_1e_2e_3$ とする．そのとき，入力文を条件として，出力文の確率を最大にするモデルのパラメータを求める．

$MaxP(e_1e_2e_3|j_1j_2j_3)$

一方 AR モデル [2],[3] は，古くから制御理論などで用いられてきたモデルである．このモデルは，現在のデータは，過去のデータの重みに雑音が重畳していると仮定している．

$X(n) = X(n-1) * a_1 + X(n-2) * a_2 + ... + X(n-m) * a_m + \varepsilon$

このモデルを encoder-decoder モデルの変形と考える．そのとき，入力データは，初期の過去データとみなすことができる．

$MaxP(e_1e_2e_3)$ ただし過去データは $j_1j_2j_3$

つまり，入力文を過去データとし，出力文の生成確率を最大にするモデルを求めることになる．

以上の考察から，本論文では，AR モデルに基づく機械翻訳を行うと仮定する．その仮定に基づいたときに，翻訳文は，尤度最大の基準で決めるのではなく．多数決で決定することを提案する．また，入力 1 文毎にモデルの最適化を行う．このとき，入力 1 文の単語数を AR モデルの回帰次数に相当すると仮定する．以上の点を考慮して翻訳実験を行う．

2 AR モデルと encoder-decoder モデル

入力文を過去のデータとみなし，出力文を AR モデルで生成すると考えれば，encoder-decoder モデルは AR モデルの 1 形態とみなすことが可能である．このモデルを図 1 にしめす．

この図では，$j_1j_2j_3$ は，過去のデータとみなす．そして $e_1e_2e_3$ を AR モデルで生成すると仮定する．

この図のように考察することで，従来の encoder-decoder モデルの学習方法において，新しい着目点が出てくる．

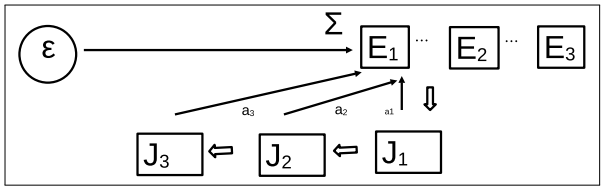

図 1 自己回帰モデルに基づく機械翻訳の概念図 （ARMT）

3 AR モデルに基づく機械翻訳

本研究において，encoder-decoder モデルは，入力文を過去データとする，AR モデルとして扱う．この提案方法を”AR モデルに基づく機械翻訳（Auto Regressive Machine Translation (ARMA)）” と呼ぶ．以下に提案手法の基本的な着目点について述べる．

3.1 多数決

AR モデルは，現在のデータは，過去のデータの重みに雑音が重畳していると仮定している．そのため，得られた尤度に信頼がない．そのため，出力文を尤度で選択する妥当性がない．

そこで，本研究では，複数の候補文における多数決を基本とする．複数のモデルを作成し，入力文に対して，複数の候補文を出力し，多数決で出力文を選択する．

3.2 未知語処理（追加データ）

AR モデルは，雑音に非常に弱いことが知られている．翻訳モデルにおいて AR モデルを想定した場合，未知語が雑音に相当する．つまり，文中に未知語があった場合，翻訳精度が大幅に減少する．

そこで，本研究では，入力文において，未知語が無いように学習文を追加する．具体的には，他の大量の対訳文から，入力文において存在する単語を含む対訳文を選択し，学習データに追加する．

3.3 入力文 1 文毎の再学習と回帰次数

AR モデルでは，回帰次数が常に問題になる．この回帰次数は，AIC[4] や BIC[5] が提案されているが，明確に妥当性のある方法がない．しかし，機械翻訳では，入力文の単語数を，回帰次数と想定できる．つまり，学習データの単語数が，入力文の単語数と同一であることに妥当性がある．そこで，入力文 1 文毎に，翻訳モデルを作成する．

4 提案方法

以下に本論文で提案する翻訳方法の概略を示す．

4.1 複数の候補文

提案手法では，1 つの入力文から，複数の候補文を出力する．そして，最終的に，多数決で，出力文を選択する．複数の候補文を得るため，本研究において採用する方法

を以下に述べる．

1. 通常 or 反転 入力 or 出力 計 4 モデル
   通常の NMT では，順方向の入力文と，順方向の出力文が想定されている．しかし，入力文の単語の順序と，出力文の単語の順序には，順方向と逆方向の組み合わせで，合計 4 つのモデルがある．以下にそれぞれのモデルを日英翻訳を前提に例を示す．
   (a) 通常入力，通常出力 （JE）
   例： 春 が 来た *Spring has come*
   (b) 反転入力，通常出力 （jE）
   例： 来た が 春 *Spring has come*
   (c) 反転入力，反転出力 （je）
   例： 来た が 春 *come has Spring*
   (d) 通常入力，反転出力 （Je）
   例： 春 が 来た *come has Spring*
2. 複数候補 N-best
   1 つのシステムにおいて複数の出力（N-best）を利用する．
3. 複数システム M-system （初期乱数）
   AR モデルでは，常に while noize が重畳していると仮定している．そのため，初期値によって結果がことなる．そこで複数の初期値において，同一のシステムを，複数構築する．
4. 入力文 1 文毎の再学習のための類似文の種類 L-corpus
   本研究では，全体のモデルを構築したあと，入力 1 文毎に，類似文を大規模な対訳データから検索して，再学習をおこなう．類似文の検索の方法の違いと，検索量から，類似文が異なるシステムを，4 種類を作成する．

4.2 未知語（追加学習）

AR モデルは，非常に高い周波数分解能を持つ．その反面，雑音に非常に弱いことが知られている．翻訳においては，未知語が雑音に相当する．未知語がない汎用的なモデルの学習は，実際には困難である．そこで，本研究では，入力文において，未知語が無いように他の大量の対訳文から文を収集する．ただし，文長には有意する（4.3 章 1 節）．

4.3 入力文 1 文毎の再学習

本研究では，入力 1 文毎に，NMT のモデルを学習する．この方式を直接行うと，計算コストが莫大になる．そこで再学習をすることで，問題の軽減をおこなう．初めに，基本学習データを利用して，基本モデルを作成する．次に入力文 1 単位に入力文の類似文を利用して基本モデルの再学習をおこなう．そのときに考慮する点を以下に示す．

1. 回帰次数と，入力文と類似文の文長
   機械翻訳において AR モデルを適応したとき，回帰次数が問題になる．この回帰次数は，明確に決定する方法がない．しかし，入力文 1 文毎に，モデルを作成することを仮定した場合，入力文の単語数が，回帰次数と想定できる．つまり，学習データの単語数が， 入力文の単語数と同一であることに妥当性がある．実際には，入力文と単語数が完全に同一である学習文は，少ないため，ある幅をもたせる．
2. 類似文検索の方法
   大量の対訳文から，入力文に類似した文を選択する．類似文の検索には，単語類似度， Levenshtein 距離，TFIDF などがある．

5 実験条件

本実験では，1 つの入力文に対し 256 文の候補文（4×4×4×4）を生成する．そして，多数決によって出力文を選択する．以下に詳細を示す．

5.1 複数システム

1. 翻訳順序 4
   提案手法では，4 つの入力出力正逆順序のモデルをつかう．（1 節）
2. 複数候補 4-best
   1 システムの入力 1 文につき，4 文の候補を出力する．
3. 初期乱数 4-system
   初期値が異なる乱数を 4 システム作成する．
4. 入力 1 文毎の学習 4-corpus
   入力文の類似文は，4 種類作成する．詳細は 5.3 章において述べる．

5.2 学習データ

1. 基本学習データ
   基本学習データは，電子辞書などから抜き出した，単文および重複文データである [6].
2. 未知語処理 （追加学習） のデータ
   未知語処理のための追加データは，基本的には，JPARACRAWL[7] から抽出する．条件を以下に示す．
   (a) 入力文 1 文毎の類似度
   入力文に対して，類似度を計算し，上位 1024 まで選択する．TF で計算する．
   (b) 入力文 1 文毎の 1 単語毎の類似度
   学習データにおいて，入力文に未知語が存在させない目的で，入力文の各単語毎に，各単語を含む文の類似度を計算し，上位 32 文を選択する．
   (c) 文の長さ
   すべての類似文は，入力文の長さの，2/3 から 3/2 までとする．

学習データ量を表 1 にまとめる．

表 1 学習データ

単文	163188 文対
複文	92427 文対
追加学習	130974 文対

なお，日本語の総語彙数は 83629 単語で， 英語の総語彙数は 92584 単語である．

5.3 入力文 1 文毎の再学習

本研究では，基本学習データを利用して，基本モデルを作成する．次に入力文 1 毎に基本モデルの再学習をおこなう．再学習には JPARACRAWL[7] から入力文毎に類似文を抽出する．類似文検索には，多くの方法があるが，本研究では，最も単純な TF と，TFIDF の 2 種類を用いる．また，類似文として上位 4 文と上位 1024 文を用いる．合計，4 種類を作成する．なお，文長には有意する（4.3 章 1 節）．

5.4 学習回数

提案手法では，全体のモデルを構築したあと，入力文 1 毎に再学習をおこなう．各学習における学習回数を表 2 に示す．

表 2 学習方法と繰り返し回数

全体のモデル	100000
1 文単位の再学習	100

5.5 テストデータ

テストデータは，電子辞書などから抜き出した，重複文データ [6] とする．基本学習データと，完全に open なデータである．評価は，100 文で行う．

5.6 最終判断

提案システムは，合計 256 文から 1 文を多数決で選択する．具体的には，複数のシステムから，同一の文の数が最も高い文を選択する．なお，同一文が複数あった場合，尤度で選択する．

5.7 翻訳システム

提案したシステムを評価するために以下の 5 システムを作成し．比較および評価を行う．以下に概要を示す．

1. ベースライン
   学習データは，基本学習データ（単文と重複文）（1 節）に，未知語処理（追加学習）のデータ（入力 100 文の類似文 5.2 章 2 節）を加える．学習データ総数は，386589 文対である．
2. 1 文毎再学習
   追加システムにおいて作成したモデルに，入力 1 文毎に，再学習をおこなう．再学習の学習データは，入力 1 文毎に類似文（5.2 章 2 節）を利用する．
3. 多数決（提案システム）
   提案システムである．1 つの再学習システムにおいて 4 候補，再学習システムを 4 組，さらに入出力反転で 4 組，文単位の学習データ 4 種類，合計 256 候補文において，多数決（5.6 章）で決定する．
4. google 翻訳
   比較および評価のため，2023 年 8 月における，google 翻訳の出力を利用する．

6 実験結果

6.1 自動評価

自動評価による実験結果を表 3 に示す．評価文は 100 文である．

表 3 自動評価

評価方法	BLEU	meteor	TER	RIBES
多数決（提案）	0.3132	0.5839	0.5039	0.8285
1 文毎再学習	0.2781	0.5444	0.5494	0.8265
ベースライン	0.2340	0.5046	0.6043	0.7959
google	0.2597	0.5535	0.5766	0.8202

以上の結果より，自動評価において，提案手法の高い性能がわかる．特に多数決をもちいたとき，翻訳の精度向上が著しい．そして google の翻訳性能を超えている．

6.2 人手評価（提案手法とベースライン）

提案手法とベースラインの対比較評価を行った．対象は約 100 文である．評価者は学生 6 名である．結果 の平均を表 4 に示す．

表 4 対比較評価（人手）

評価	割合	件数
提案手法 > ベースライン	25.1%	151
提案手法 < ベースライン	11.6%	70
提案手法 ベースライン 両者 ○	40.8%	245
提案手法 ベースライン 両者 ×	22.3%	134

以上の結果より，人手評価でも，提案手法はベースラインを大幅に超えていることがわかる．

6.3 出力例

翻訳例を表 5 に示す．

表 5 翻訳例
多数決の（）内の数字は，同一の翻訳文の数

テスト文 1	その 足音 が いっそう 近づいた か と 思う と また 遠ざかった
参照文	The footsteps came nearer , then went away again .
google	The footsteps moved closer and further away.
ベースライン	The sound of the footsteps drew over again .
1 文毎再学習	The footsteps swam over again .
多数決 (3)	The footsteps approached and then fell away .
テスト文 1	細君 は 派手 好きで 金 さえ あれば 着物 を 買って しまう
参照文	His wife is a dashing woman , and spends all her money on finery .
google	My wife likes gaudiness and will buy a kimono if she has the money.
ベースライン	His wife is very fond of her finery and buys her clothes .
1 文毎再学習	His wife is fond of finery , and spends all his money on clothes .
多数決 (22)	His wife is fond of finery , and spends all his money on clothes .

7 考察

7.1 翻訳失敗の原因

翻訳が誤った文を調査した．その結果，多くの文は，未知語と類似文の問題であった．例をあげて述べる

1. 未知語
   学習文に存在しない単語が出現する．
   例” 村人 に 危険 を 知らせる ため に 早鐘 を ついた “
   “早鐘” が学習データに存在しない．
2. 類似文なし
   例 “みだりに 干渉 して は 有害 無益だ ”
   この例では，類似する文が存在しなかった．最も類似した文は以下の文であった
   ” 緊張 して は いけ ませ ん 。” “Do not be nervous . “
3. 類似文あり．単語の類似文なし
   例 “私 は 針 の むしろ に 座る 心地 が した “
   類似文は存在している．
   “針 の むしろ に 座って いる ような 心境 だった ”
   “I felt as if I were sitting on thorns . “
   しかし，“心地 “を含む学習データが存在しない．JPARACRAWL では “心地” を含む対訳文対は，存在している．しかし，いずれも入力文に類似していないため，追加学習から削除されている．
   “冷たい 空気 が ほろ酔い の 頬 に 触れる 時 の 感覚 も 心地 良い ”
   “I like the feeling of cold air touching my cheeks when I’m tipsy .”
   “遊ぶ だけ 遊ば ない と 飽き足ら ぬ 心地 が する ”
   “If you do not take your fill of pleasure , you have an unsatisfied

feeling . “

纏めると，翻訳が誤る原因は，以下の 2 つの条件の and 条件であった．

1. 類似文がない．
2. 未知語がある．

逆に，この 2 点がない文は，基本的には正しい翻訳が得られた．これらの問題の解決方法は，学習データを増加するしかないと考えている．

7.2 最終判断の方法

提案システムは，合計 256 文から 1 文を選択する．この選択方法は，以下の 3 種類がある．

1. 多数決
   複数のシステムから，多数決で選択する．候補文が複数存在した場合，尤度の総和で選択する．
2. 尤度加算
   複数システムにおいて同じ候補があった場合，その尤度を加算する．そのため，多数決に近い候補が選択される．
3. 尤度最大
   複数のシステムから尤度最大の候補文を選択する．

結果を以下に示す．

表 6 選択基準を変えた実験

選択基準	BLEU	meteor	TER	RIBES
多数決（提案）	0.3132	0.5839	0.5039	0.8285
尤度加算	0.3017	0.5729	0.5144	0.8034
尤度最大	0.2830	0.5493	0.5267	0.8058

以上の結果から，多数決が最も良いことがわかる．そして，従来の尤度最大基準は，翻訳精度が低いことがわかる．

7.3 1 文毎の再学習の効果

本研究では，全体のモデルを構築したあと，入力 1 文毎に，入力文に対する類似文を大規模な対訳データから検索して，再学習をおこなう．それぞれの 1 文毎の学習文数と検索方法と，多数決処理を行ったあとの翻訳性能を表 7 に示す．

表 7 1 文毎の学習における類似文の数と翻訳率の変化

検索方法	類似文の数	BLEU	meteor	TER	RIBES	学習回数
TF-IDF	8	0.2978	0.5733	0.5083	0.8300	100
TF	4	0.3026	0.5661	0.5179	0.8172	100
TF	64	0.2965	0.5626	0.5179	0.8060	100
TF	1024	0.2887	0.5792	0.5127	0.8282	1000

この表をみると，類似文が 4 文でも，高い翻訳精度を得ていることがわかる．また，検索方法には，大きな差がないこともわかる．これは，1 文毎の再学習が高い翻訳精度を得られることも示している．なお，これは類推に基づく翻訳 [8] によく似た形になっている．

7.4 人手評価 (提案手法と google)

提案手法と google の対比較評価を行った．対象は約 100 文である．評価者は学生 8 名である．結果 の平均を表 8 に示す．

表 8 対比較評価（人手）

評価	割合	件数
提案手法 > google	10.5%	84
提案手法 < google	36.3%	290
提案手法 google 両者 ○	44.3%	354
提案手法 google 両者 ×	9%	72

以上の結果より，人手評価では，提案手法は google に及ばないことがわかる．

7.5 ReLU と線形回帰分析とニューラルネットワーク

AR モデルは，本来，線形回帰モデルである．また，“ReLU” は，閾値をもつ線形関数とみなせる．したがって “ReLU” は，AR モデルに親和性が高い．ニューラルネットワークの学習には，従来 Sigmoid 関数や tanh 関数が使われてきた．しかし，“ReLU” を利用した学習は，ニューラルネットワークと言えるのだろうか？著者は従来の古典的な線形回帰分析と考えている．

8 おわりに

本研究では，自己回帰モデルに基づく機械翻訳 (ARMT) を提案した．このモデルでは，入力文を過去データとみなし，出力文（翻訳文）を生成するモデルになる．そして，選択基準として，尤度の代わりに，多数決を利用した．また，入力文の単語数を回帰次数と仮定し，入力文の類似文を利用して，入力文毎に NMT のモデルを再学習する．また，線形回帰分析を基本概念とした．

その結果，ベースラインと比較して大幅な性能向上が見られた．また，約 50 万文対の学習データにも関わらず，自動評価では google 翻訳を超えた．ただし人手評価では google 翻訳に及ばない．翻訳精度が低い文を調査すると，原因は，未知語と入力文の類似文の不足にあることが示された．この結果は，多数決による候補文の選択と，文単位での学習の有効性を示している．そして，自己回帰モデルに基づく機械翻訳の有効性を示している．

今後，人手評価においても，google 翻訳を超える性能を目指したい．なお，入力データを過去データとし，翻訳文を自己回帰モデルの出力データとする考え方は，対話や言い換えや要約などの多くの自然言語の分野に応用できる．今後，これらの応用も追求していきたい．

謝辞

人手評価に参加した以下の方々に感謝します．柳原　弘哉, 三木 謙志, 名村 太一, 丸山 京祐, 松本 武尊, 宮本 歩, 細川 楓, 駿河 樹

参考文献

[1] Sequence to sequence learning with neural networks. **Computation and Language**.
[2] G. Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer’s sunspot numbers.
[3] Gilbert Walker. On periodicity in series of related terms.
[4] H. Akaike. Information theory and an extension of the maximum likelihood principle. **Proceedings of the 2nd International Symposium on Information Theory**, pp. 267–281, 1973.
[5] Gideon E. Schwarz. Estimating the dimension of a model. **Annals of Statistics**, Vol. 6, p. 461–464, 1978.
[6] 村上仁一, 藤波進. 日本語と英語の対訳文対の収集と著作権の考察. 第一回コーパス日本語学ワークショップ, pp. 119–130, 2012.
[7] 森下他. Jparacrawl v3.0: 大規模日英対訳コーパス. 言語処理学会第 28 回年次大会, 2022.
[8] 村上仁一. 相対的意味論と機械翻訳の応用. 言語処理学会第 24 回年次大会, No. E5-3, pp. 924–927, 2021.