論理的アプローチと論理的アプローチを同時に行うこと によるニューラル・マシン翻訳

TRANSLATE

Dzmitry Bahdanau
Jacobs University Bremen, Germany

KyungHyun Cho **Yoshua Bengio***
Université de Montréal

ABSTRACT

ニューラル機械翻訳は、最近提案された機械翻訳のアプローチである。従来の統計的機械翻訳とは異なり、ニューラル機械翻訳は、翻訳性能を最大化するために共同調整可能な単一のニューラルネットワークを構築することを目的としている。最近提案されたニューラル機械翻訳のモデルは、エンコーダ・デコーダのファミリーに属し、原文を固定長のベクトルにエンコードし、そこからデコーダが翻訳を生成することが多い。本論文では、固定長ベクトルの使用がこの基本的なエンコーダ・デコーダアーキテクチャの性能を向上させるボトルネックであると推測し、これらの部分を明示的にハードセグメントとして形成することなく、モデルがターゲット単語の予測に関連するソース文の部分を自動的に(ソフト)探索できるようにすることでこれを拡張することを提案する。この新しいアプローチにより、英語からフランス語への翻訳タスクにおいて、既存の最先端フレーズベースのシステムに匹敵する翻訳性能を達成した。さらに、定性的な分析により、モデルによって見出された(ソフト)アライメントは、我々の直感とよく一致することが明らかになった。

1 INTRODUCTION

ニューラル機械翻訳は、最近Kalchbrenner and Blunsom (2013), Sutskever et al. (2014) and Cho et al. (2014b) によって提案された、機械翻訳への新しいアプローチである。従来のフレーズベース翻訳システム(例えば、Koehn et al., 2003参照)が、個別に調整された多数の小さなサブコンポーネントから構成されているのとは異なり、ニューラル機械翻訳は、文章を読み、正しい翻訳を出力する単一の大きなニューラルネットワークを構築し、訓練しようとするものである。

提案されたニューラル機械翻訳モデルのほとんどは、エンコーダ-デコーダのファミリー(Sutskever et al., 2014; Cho et al., 2014a)に属し、各言語のエンコーダとデコーダを持つか、各文に言語固有のエンコーダを適用し、その出力を比較する(Hermann and Blunsom, 2014)。エンコーダーニューラルネットワークは、原文を読み込んで固定長のベクトルにエンコードする。次にデコーダが符号化されたベクトルから翻訳を出力する。言語ペアのエンコーダとデコーダからなるエンコーダ・デコーダシステム全体は、原文が与えられたときに正しい翻訳が行われる確率を最大化するように共同で学習される。

このエンコーダ・デコーダのアプローチの潜在的な問題は、ニューラルネットワークが原文の必要な情報をすべて固定長のベクトルに圧縮できる必要があることである。このため、ニューラルネットワークが長い文章、特に学習コーパスの文章より長い文章に対応することが難しくなる可能性がある。Choら(2014b)は、入力文の長さが長くなるにつれて、基本的なエンコーダ・デコーダの性能が急速に劣化することを示した。

この問題に対処するために、整列と翻訳を共同で学習するエンコーダ・デコーダモデルの拡張を導入する。提案モデルは、翻訳中の単語を生成するたびに、最も関連性の高い情報が集中している原文中の位置の集合を(ソフト)検索する。次に、これらのソース位置と以前に生成されたすべてのターゲット単語に関連するコンテキストベクトルに基づいて、ターゲット単語を予測するモデルである。

*CIFAR Senior Fellow

このアプローチの基本的なエンコーダ・デコーダとの最も重要な特徴は、入力文全体を単一の固定長ベクトルにエンコードしようとしないことである。代わりに、入力文を一連のベクトルにエンコードし、翻訳をデコードしながら、これらのベクトルのサブセットを適応的に選択する。これにより、ニューラル翻訳モデルは、原文の長さに関係なく、すべての情報を固定長のベクトルにつぶす必要がなくなる。これにより、モデルが長い文章にうまく対処できることを示す。

本論文では、整列と翻訳を共同で学習する提案アプローチが、基本的なエンコーダ・デコーダアプローチよりも大幅に翻訳性能を向上させることを示す。この改善は長い文ほど顕著であるが、どのような長さの文でも観察できる。英語からフランス語への翻訳タスクにおいて、提案アプローチは単一のモデルで、従来のフレーズベースのシステムに匹敵するか、それに近い翻訳性能を達成する。さらに、定性的な分析により、提案モデルは原文と対応する目的文の間に言語的にもっともらしい(ソフトな)整合を見出すことが明らかになった。

2 背景:ニューラル・マシン翻訳

確率論的な観点からは、翻訳は原文xが与えられたときにyの条件付き確率を最大化する目的文yを見つけること、すなわち$\arg\max_y p(y \mid x)$と等価である。ニューラル機械翻訳では、並列学習コーパスを用いて、文対の条件付き確率を最大化するパラメータ化モデルを当てはめる。条件付き分布が翻訳モデルによって学習されると、原文が与えられると、条件付き確率を最大化する文を検索することによって、対応する翻訳を生成することができる。

最近、この条件分布を直接学習するためにニューラルネットワークを使用することを提案する論文が多数ある(例えば、Kalchbrenner and Blunsom, 2013; Cho et al., 2014a; Sutskever et al., 2014; Cho et al., 2014b ; Forcada and Ñeco, 1997を参照)。このニューラル機械翻訳アプローチは通常2つのコンポーネントから構成され、最初のコンポーネントは原文xをエンコードし、2番目のコンポーネントは目的文yにデコードする。例えば、(Cho et al., 2014a)と(Sutskever et al., 2014)では、2つのリカレントニューラルネットワーク(RNN)を用いて、可変長の原文を固定長のベクトルにエンコードし、そのベクトルを可変長の目標文にデコードしている。

ニューラル機械翻訳は非常に新しいアプローチであるにもかかわらず、すでに有望な結果を示している。Sutskeverら(2014)は、長短期記憶(LSTM)ユニットを持つRNNに基づくニューラル機械翻訳が、英語からフランス語への翻訳タスクにおいて、従来のフレーズベース機械翻訳システムの最先端性能に近い性能を達成することを報告した[1] 既存の翻訳システムにニューラルコンポーネントを追加する、例えば、フレーズテーブルのフレーズペアをスコアリングしたり(Cho et al, 2014a)、翻訳候補を再ランク付けしたり(Sutskever et al, 2014)することで、p最先端の性能レベルを達成した。

2.1 RNN符号化器-符号化器

ここでは、Choら(2014a)とSutskeverら(2014)によって提案されたRNN Encoder-Decoderと呼ばれる基礎となるフレームワークについて簡単に説明する。

エンコーダ・デコーダの枠組みでは、エンコーダは入力文、ベクトル列 $x = (x_1, --, x_{Tx})$ をベクトル c に読み込む。

$$h_t = f(x_t, h_{t-1}) \tag{1}$$

and

$$c = q(\{h_1, \cdots, h_{T_x}\}),$$

ここで、$h_t \in R^n$は時刻tにおける隠れ状態、cは隠れ状態のシーケンスから生成されるベクトルである。fとqはいくつかの非線形関数である。Sutskeverら(2014)は、例えばfとq $(\{h_1, --, h_T\}) = h_T$としてLSTMを用いた。

[1] 最先端の性能とは、ニューラルネットワークベースのコンポーネントを使用しない、従来のフレーズベースシステムの性能を意味する。

[2] 先行研究(例えば、Cho et al., 2014a; Sutskever et al., 2014; Kalchbrenner and Blunsom, 2013参照)の多くは可変長の入力文を固定長ベクトルにエンコードするために用いたが、後で示すように、可変長ベクトルを持つことは必要ではなく、有益である可能性さえある。

デコーダは、文脈ベクトルcと以前に予測された全ての単語{y_1, – –, $y_{t\ 0-1}$ }が与えられたとき、次の単語$y_{t\ 0}$を予測するように学習されることが多い。つまり、デコーダは、結合確率を順序付き条件式に分解することで、翻訳yに対する確率を定義する。

$$p(\mathbf{y}) = \prod_{t=1}^{T} p(y_t \mid \{y_1, \cdots, y_{t-1}\}, c), \tag{2}$$

ここで、y = y_1, – –, y_{Ty}.RNNでは、各条件付き確率は次のようにモデル化される。

$$p(y_t \mid \{y_1, \cdots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c), \tag{3}$$

ここで、gはy_tの確率を出力する非線形、潜在的に多層化された関数であり、s_tはRNNの隠れ状態である。RNNと脱畳み込みニューラルネットワークのハイブリッドなど、他のアーキテクチャも使用できることに注意すべきである(Kalchbrenner and Blunsom, 2013)。

3 L3 整列と翻訳への獲得

本節では、ニューラル機械翻訳のための新しいアーキテクチャを提案する。新しいアーキテクチャは、エンコーダとしての双方向RNN(第3.2節)と、翻訳をデコードする際に原文を検索することをエミュレートするデコーダ(第3.1節)から構成される。

3.1 DECODER : 一般的な説明

新しいモデル・アーキテクチャでは、式(2)の各条件付き確率を次のように定義する:

$$p(y_i|y_1, \ldots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i), \tag{4}$$

ここで、s_iは時間iのRNN隠れ状態であり、次式で計算される。

$$s_i = f(s_{i-1}, y_{i-1}, c_i).$$

既存のエンコーダ・デコーダのアプローチ(式(2)参照)とは異なり、ここでは確率は各ターゲット単語y_iに対して異なる文脈ベクトルc_iに条件付けされることに注意すべきである。

文脈ベクトルc_iは、エンコーダが入力文をマッピングする一連の注釈(h_1, – –, h_{Tx})に依存する。各注釈h_iは、入力シーケンスのi番目の単語を取り囲む部分に強く焦点を当てた、入力シーケンス全体に関する情報を含む。アノテーションの計算方法については、次のセクションで詳しく説明する。

文脈ベクトルc_iは、これらの注釈h_iの加重和として計算される:

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j. \tag{5}$$

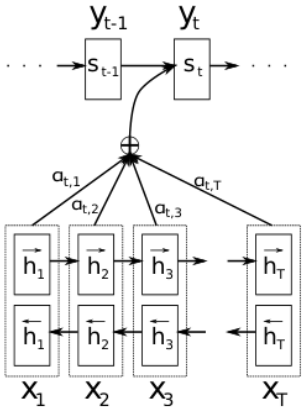

図1:原文(x_1, x_2,.., x_T)が与えられたときに、t番目のターゲット単語y_tを生成しようとする提案モデルの図解。

各注釈h_jの重みα_{ij}は次式で計算される。

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}, \tag{6}$$

where

$$e_{ij} = a(s_{i-1}, h_j)$$

は位置j周辺の入力と位置iの出力がどの程度一致するかをスコア化するアライメントモデルである。スコアはRNNの隠れ状態s_{i-1}(y_iを放出する直前、式(4))と入力文のj番目のアノテーションh_jに基づいている。

アライメントモデルaを、提案システムの他のすべてのコンポーネントと共同で学習されるフィードフォワードニューラルネットワークとしてパラメトリック化する。

従来の機械翻訳とは異なり、アライメントは潜在変数とはみなされないことに注意してください。代わりに、アライメントモデルはソフトアライメントを直接計算し、コスト関数の勾配をバックプロパゲートすることができる。この勾配は、アライメントモデルだけでなく、翻訳モデル全体の共同学習にも利用できる。

すべてのアノテーションの加重和を期待されるアノテーションの計算として取るというアプローチを理解することができる。α_{ij}を、対象単語y_iが原語x_jに整列、または原語から翻訳される確率とする。そして、i番目のコンテキストベクトルc_iは、確率α_{ij}で全てのアノテーションに対する期待アノテーションである。

確率α_{ij}、またはそれに関連するエネルギーe_{ij}は、次の状態s_iを決定しy_iを生成する際の、前の隠れ状態s_{i-1}に関する注釈h_jの重要性を反映する。直感的には、これはデコーダに注意のメカニズムを実装している。デコーダは、ソース文の一部に注意を払うように決定する。デコーダに注意機構を持つことで、原文中のすべての情報を固定長のベクトルにエンコードしなければならないという負担からエンコーダを解放することができる。この新しいアプローチでは、情報をアノテーションのシーケンス全体に分散させることができ、それに応じてデコーダが選択的に取り出すことができる。

3.2 エンコーダ:シーケンスを注釈するための双方向RNN

式(1)で記述される通常のRNNは、入力シーケンスxを最初のシンボルx_1から最後のシンボルx_{T_x}まで順番に読み込む。しかし、提案方式では、各単語のアノテーションは、先行する単語だけでなく、以下の単語も要約するようにしたい。そこで、最近音声認識でうまく利用されている双方向RNN(BiRNN, Schuster and Paliwal, 1997)を用いることを提案する(例えば、Graves et al, 2013を参照)。

→ – BiRNNは前方RNNと後方RNNから構成される。前方RNN fは入力列→→→→→→→→→→→→→→→→→→→→→→→→→→→→→
→→
→→
→→
→→→→→→→→→→→→→→→→→→→→→→→→→→→→→→← – 後方RNN fは(x_{T_x}からx_1までの)逆順にシーケンスを読み、その結果、後方隠れ状態(h_1, – –, h_{T_x})のシーケンスが← – ← – となる。

→ – 各単語x_jに対して、前方隠れ状態h_jと

the

j j j j j は、先行する単語と後続する単語の両方である。RNNは最近の入力をよりよく表現する傾向があるため、アノテーションh_jはx_jの周りの単語に焦点を当てる。この一連の注釈は、デコーダとアライメントモデルが後にコンテキストベクトルを計算するために使用する(式(8))。(5)-(6)).

提案モデルの図解は図1を参照のこと。

4 実験セットアップ

提案手法を英仏翻訳のタスクで評価する。比較として、Choら(2014a)が最近提案したRNN Encoder-Decoderの性能も報告する。両モデルとも同じ学習手順と同じデータセットを使用する[4]。

4.1 Dataset

WMT '14には、以下の英仏並列コーパスが含まれている:Europarl(6100万語)、ニュース解説(550万語)、UN(421万語)、クロールされた2つのコーパス(それぞれ9000万語と2725万語)、合計8億5000万語。Choら(2014a)に記載された手順に従い、Axelrodら(2011)によるデータ選択法を用いて、結合コーパスのサイズを348M語に縮小する。[5] エンコーダの事前学習にもっと大きな単言語コーパスを用いることも可能かもしれないが、言及した並列コーパス以外の単言語データは用いない。ニューステストを連結する

[3] http://www.statmt.org/wmt14/translation-task.html
[4] Implementations are available at https://github.com/lisa-groundhog/GroundHog.
[5] Available online at http://www-lium.univ-lemans.fr/~schwenk/cslm_joint_paper/.

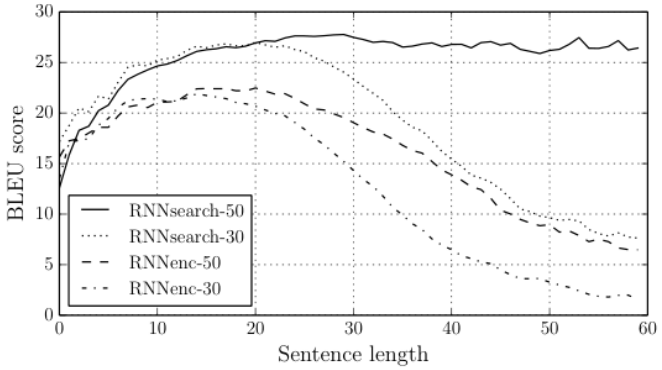

図2:テストセットで生成された翻訳文の文の長さに対するBLEUスコア。結果は、モデルに未知の単語を含む完全なテストセットでのものである。

2012年とnews-test-2013で開発(検証)セットを作成し、学習データに存在しない3003文からなるWMT '14のテストセット(news-test-2014)でモデルを評価する。

通常のトークン化[6]の後、各言語で最も頻度の高い単語30,000語のショートリストを用いてモデルを学習する。ショートリストにない単語は特別なトークン([UNK])にマップされる。データには小文字やステミングなど、その他の特別な前処理は適用しない。

4.2 MODELS

我々は2種類のモデルを訓練する。1つ目はRNN Encoder-Decoder (RNNencdec, Cho et al., 2014a)であり、もう1つは提案モデルである。各モデルを2回訓練する:1回目は30語までの文(RNNencdec-30、RNNsearch-30)、次に50語までの文(RNNencdec-50、RNNsearch-50)である。

RNNencdecのエンコーダとデコーダはそれぞれ1000個の隠れユニットを持つ。[7] RNNsearchのエンコーダは、それぞれ1000個の隠れユニットを持つ前方リカレントニューラルネットワーク(RNN)と後方リカレントニューラルネットワーク(RNN)から構成される。そのデコーダは1000個の隠れユニットを持つ。どちらの場合も、各ターゲット単語の条件付き確率を計算するために、単一のマックスアウト(Goodfellow et al., 2013)隠れ層を持つ多層ネットワークを使用する(Pascanu et al., 2014)。

ミニバッチ確率的勾配降下(SGD)アルゴリズムとAdadelta(Zeiler, 2012)を用いて各モデルを学習する。各SGD更新方向は、80文のミニバッチを用いて計算される。各モデルを約5日間学習させた。

モデルが学習されると、ビームサーチを用いて、条件付き確率を近似的に最大化する翻訳を見つける(例えば、Graves, 2012; Boulanger-Lewandowski et al.)Sutskeverら(2014)は、このアプローチを用いて、ニューラル機械翻訳モデルから翻訳を生成した。

実験に使用したモデルのアーキテクチャと学習手順の詳細については、付録AおよびBを参照のこと。

5 RESULTS

5.1 定量的結果

表1に、BLEUスコアで測定された翻訳性能を示す。表から明らかなように、全てのケースにおいて、提案するRNNsearchは従来のRNNencdecを凌駕している。さらに重要なことは、既知の単語からなる文のみを考慮した場合、RNNsearchの性能は従来のフレーズベース翻訳システム(Moses)と同程度に高いということである。これは、MosesがRNNsearchとRNNencdecの学習に使用した並列コーパスに加えて、別のモノリンガルコーパス(418M語)を使用していることを考慮すると、重要な成果である。

[6] オープンソースの機械翻訳パッケージMosesのトークン化スクリプトを使用した。本稿では、「隠れユニット」とは、常にゲート付き隠れユニットを意味する(付録A.1.1参照)。

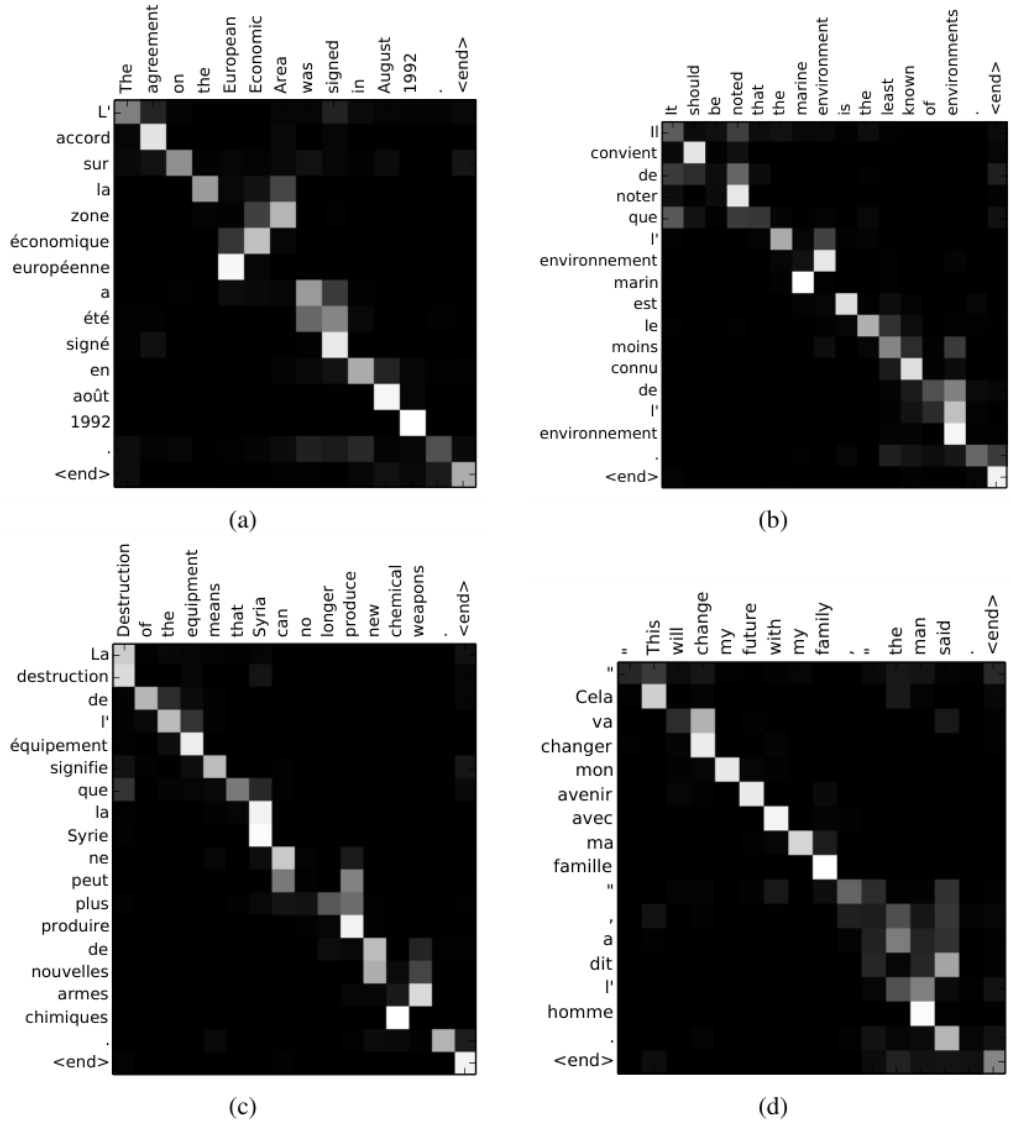

図3:RNNsearch-50によって発見された4つのアライメント例。各プロットのx軸とy軸は、それぞれ原文(英語)と生成された翻訳(フランス語)の単語に対応する。各ピクセルは、i番目のターゲット単語に対するj番目のソース単語のアノテーションの重みα_{ij}をグレースケール(0:黒、1:白)で示す(式(6)参照)。(a) 任意の文。(b-d)テストセットから、未知の単語がなく、長さが10～20語の文の中からランダムに選んだ3つのサンプル。

提案されたアプローチの動機の1つは、基本的なエンコーダ・デコーダのアプローチで固定長のコンテキストベクトルを使用することであった。この制限により、基本的なエンコーダ・デコーダのアプローチは、長い文に対して性能が低下する可能性があると推測した。図2より、RNNencdecの性能は文の長さが長くなるにつれて劇的に低下することがわかる。一方、RNNsearch-30とRNNsearch-50は文の長さに対してより頑健である。RNNsearch50は、特に、長さ50以上の文でも性能劣化が見られない。この基本的なエンコーダ・デコーダに対する提案モデルの優位性は、RNNsearch-30がRNNencdec-50を上回ることさえある(表1参照)。

Model	All	No UNK°
RNNencdec-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNencdec-50	17.82	26.71
RNNsearch-50	26.75	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

表1:テストセットで計算された学習済みモデルのBLEUスコア。2列目と3列目は、それぞれすべての文と、それ自体と参照訳に未知の単語がない文のスコアを示している。RNNsearch-50?は開発セットでの性能が向上しなくなるまでずっと長く訓練されたことに注意。(*) 未知の単語を持たない文のみを評価した場合、[UNK]トークンを生成することをモデルに禁止した(最後の列)。

5.2 定性的分析

5.2.1 整列

提案手法は、生成された翻訳文の単語と原文の単語との間の(ソフト)アライメントを直感的に検査する方法を提供する。これは、図3のように式(6)からアノテーションの重みα_{ij}を可視化することで行われる。各プロットの行列の各行は、アノテーションに関連する重みを示す。このことから、原文のどの位置が、目的語を生成する際に重要であると考えられるかがわかる。

図3のアライメントから、英語とフランス語の単語のアライメントはほぼ単調であることがわかる。各行列の対角線に沿って強い重みが見られる。しかし、自明でない、単調でないアライメントも多数観察される。形容詞と名詞は通常、フランス語と英語では異なる順序で並べられ、図3(a)にその例を示す。この図から、このモデルは[欧州経済地域]というフレーズを[zone économique européen]に正しく翻訳していることがわかる。RNNsearchは[zone]を[Area]と正しく整列させ、2つの単語([European]と[Economic])を飛び越え、一度に1単語を振り返ってフレーズ全体[zone économique européenne]を完成させることができた。

ハードアライメントとは対照的に、ソフトアライメントの強さは、例えば図3(d)から明らかである。l' homme]に翻訳された原語句[the man]を考えてみよう。ハードアライメントは[the]を[l']に、[man]を[homme]に写像する。これは翻訳には役立たない。[le]、[la]、[les]、[l']のどちらに翻訳すべきかを判断するために、[the]に続く単語を考えなければならないからである。我々のソフトアライメントは、モデルに[the]と[man]の両方を見てさせることで、この問題を自然に解決し、この例では、モデルが[the]を[l']に正しく変換できたことがわかる。図3では、すべてのケースで同様の動作が観察される。ソフトアライメントのさらなる利点は、いくつかの単語をどこにもマッピングしたり、どこからマッピングしたりする直感に反する方法([NULL])を必要とせず、異なる長さのソースフレーズとターゲットフレーズを自然に扱うことである(例えば、Koehn, 2010の第4章と第5章を参照)。

5.2.2 長期的な出来事

図2から明らかなように、提案モデル(RNNsearch)は、長文の翻訳において、従来のモデル(RNNencdec)よりもはるかに優れている。これは、RNNsearchが長い文を固定長のベクトルに完全にエンコードする必要はなく、入力文のうち特定の単語を取り囲む部分のみを正確にエンコードするという事実によるものと思われる。

例として、テストセットからこの原文を考えてみよう:

> 入院特権とは、病院における医療従事者としての地位に基づき、患者を病院または医療センターに入院させ、診断または処置を実施する権利のことである。

RNNencdec-50はこの文章を次のように翻訳した:

> 入院の特権とは、患者を病院または医療センターに認定する、または診断を健康状態に応じて行う、医師の権利である。

RNNencdec-50は、[医療センター]まで原文を正しく翻訳した。しかし、そこから(下線部)、原文の本来の意味から逸脱していた。例えば、原文の[病院における医療従事者としての地位に基づく]を[en fonction de son état de santé]に置き換えた(「健康状態に基づく」)。

一方、RNNsearch-50は、入力文の意味をすべて保持したまま、細部を省略することなく、次のような正しい訳語を生成した:

> 入院の特権とは、病院または医療センターに患者を入院させ、病院への健康管理の業務慣行に従って、診断または処置を行う権利を有する医師である。

テストセットから別の文を考えてみよう:

> このような経験は、ディズニーが「シリーズの寿命を延ばし、デジタル・プラットフォームを通じてオーディエンスと新たな関係を構築し、ますます重要性を増している」という取り組みの一環である。

RNNencdec-50による翻訳は

> この種の経験は、ディズニーの「新しい生活の持続期間を延長し、より複雑になるデジタル読者とのつながりを発展させる」というイニシアチブの一部である。

前の例と同様に、RNNencdecは約30語(下線部参照)を生成した後、原文の実際の意味から逸脱し始めた。その後、翻訳の質は低下し、引用符が閉じていないなどの基本的な間違いが生じる。

ここでも、RNNsearch-50はこの長い文章を正しく翻訳することができました:

> この種の経験は、ディズニーが「シリーズの存続期間を延長し、より重要なデジタルプラットフォームを通じて、一般の人々との新しい関係を構築する」という努力の一環であった。

すでに示した定量的な結果と合わせて、これらの定性的な観察から、RNNsearchアーキテクチャは標準的なRNNencdecモデルよりもはるかに信頼性の高い長文の翻訳を可能にするという我々の仮説が確認された。

付録Cでは、RNNencdec-50、RNNsearch-50、Google翻訳によって生成された長文原文の翻訳例を、参照訳とともにさらにいくつか示す。

6 関連研究

6.1 整列への獲得

出力記号を入力記号と整合させる同様のアプローチは、最近Graves(2013)によって手書き合成の文脈で提案された。手書き合成は、モデルが与えられた一連の文字の手書きを生成するように要求されるタスクである。彼の研究では、アノテーションの重みを計算するためにガウスカーネルの混合を使用し、各カーネルの位置、幅、混合係数はアライメントモデルから予測された。より具体的には、彼のアライメントは、位置が単調に増加するような位置を予測するように制限されていた。

我々のアプローチとの主な違いは、(Graves, 2013)では、アノテーションの重みのモードが一方向にしか動かないことである。機械翻訳の文脈では、文法的に正しい翻訳(例えば英語からドイツ語)を生成するために、(長距離の)並べ替えが必要になることが多いため、これは深刻な制限である。

一方、我々のアプローチでは、翻訳中の各単語について、原文中の各単語のアノテーション重みを計算する必要がある。この欠点は、入力文と出力文のほとんどが15-40語しかない翻訳のタスクでは深刻ではない。しかし、このため、提案方式の他のタスクへの適用が制限される可能性がある。

6.2 機械翻訳のためのニューラルネットワーク

Bengioら(2003)がニューラルネットワークを用いて、先行する単語の固定数が与えられたときの単語の条件付き確率をモデル化するニューラル確率的言語モデルを導入して以来、ニューラルネットワークは機械翻訳で広く用いられてきた。しかし、ニューラルネットワークの役割は、既存の統計的機械翻訳システムに単一の特徴を提供するか、既存のシステムによって提供された翻訳候補のリストを再ランク付けすることだけに大きく制限されてきた。

例えば、Schwenk (2012)は、フィードフォワードニューラルネットワークを使用して、ソースフレーズとターゲットフレーズのペアのスコアを計算し、そのスコアをフレーズベースの統計的機械翻訳システムの追加機能として使用することを提案した。より最近では、Kalchbrenner and Blunsom (2013)とDevlin et al. (2014)が、既存の翻訳システムのサブコンポーネントとしてニューラルネットワークの成功例を報告している。従来、ターゲット側言語モデルとして学習されたニューラルネットワークは、翻訳候補のリストの再スコアや再ランク付けに使用されてきた(例えば、Schwenk et al.)

上記のアプローチは、最先端の機械翻訳システムよりも翻訳性能を向上させることが示されたが、我々は、ニューラルネットワークに基づく全く新しい翻訳システムを設計するという、より野心的な目的により興味を持っている。したがって、本稿で検討するニューラル機械翻訳アプローチは、これらの先行研究とは根本的に異なるものである。既存のシステムの一部としてニューラルネットワークを使用するのではなく、我々のモデルは単独で動作し、原文から直接翻訳を生成する。

7 Conclusion

従来のニューラル機械翻訳のアプローチは、エンコーダ・デコーダ・アプローチと呼ばれ、入力文全体を固定長のベクトルにエンコードし、そこから翻訳がデコードされる。Choら(2014b)とPouget-Abadieら(2014)が報告した最近の実証研究に基づき、固定長の文脈ベクトルの使用は長文の翻訳に問題があると推測した。

本論文では、この問題に対処するための新しいアーキテクチャを提案した。各ターゲット単語を生成する際に、入力単語の集合、またはエンコーダによって計算されたそれらの注釈をモデルに(ソフト)探索させることで、基本的なエンコーダ・デコーダを拡張した。これにより、モデルは原文全体を固定長のベクトルにエンコードする必要がなくなり、次の目的語の生成に関連する情報のみに焦点を当てることができる。これは、ニューラル機械翻訳システムがより長い文章に対して良い結果をもたらす能力に大きなプラスの影響を与える。従来の機械翻訳システムとは異なり、アライメント機構を含む翻訳システムのすべての部分は、正しい翻訳を生成するより良い対数確率に向けて共同で学習される。

RNNsearchと呼ばれる提案モデルを、英仏翻訳のタスクでテストした。実験の結果、提案するRNNsearchは、文の長さに関係なく、従来のエンコーダ・デコーダモデル(RNNencdec)を大幅に上回り、原文の長さに対してより頑健であることが明らかになった。RNNsearchによって生成された(ソフト)アライメントを調査した定性的な分析から、モデルは正しい翻訳を生成するため、各ターゲット単語をソース文の関連単語、またはその注釈と正しくアライメントできると結論付けることができた。

おそらくより重要なことは、提案されたアプローチは、既存のフレーズベースの統計的機械翻訳に匹敵する翻訳性能を達成したことである。提案されたアーキテクチャ、あるいはニューラル機械翻訳の全ファミリーが今年になってようやく提案されたばかりであることを考えると、これは驚くべき結果である。ここで提案するアーキテクチャは、より良い機械翻訳と自然言語全般のより良い理解への有望な一歩であると考える。

今後の課題として残されているのは、未知語や希少語をよりよく扱うことである。これは、モデルがより広く使用され、すべてのコンテキストで現在の最先端の機械翻訳システムの性能に匹敵するために必要である。

ACKNOWLEDGMENTS

Theano (Bergstra et al., 2010; Bastien et al., 2012)の開発者に感謝したい。研究資金および計算機サポートについて、以下の機関の支援に感謝する：NSERC、Calcul Québec、Compute Canada、Canada Research Chairs、CIFAR。BahdanauはPlanet Intelligent Systems GmbHの支援に感謝する。また、フェリックス・ヒル、バート・ヴァン・メリエンボア、ジャン・プジェ=アバディ、コリン・デビン、キム・テホにも感謝する。

REFERENCES

Axelrod, A., He, X., and Gao, J. (2011). Domain adaptation via pseudo in-domain data selection. In *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 355–362. Association for Computational Linguistics.

Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I. J., Bergeron, A., Bouchard, N., and Bengio, Y. (2012). Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, **5**(2), 157–166.

Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.*, **3**, 1137–1155.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., and Bengio, Y. (2010). Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*. Oral Presentation.

Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P. (2013). Audio chord recognition with recurrent neural networks. In *ISMIR*.

Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., and Bengio, Y. (2014a). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*. to appear.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014b). On the properties of neural machine translation: Encoder–Decoder approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. to appear.

Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., and Makhoul, J. (2014). Fast and robust neural network joint models for statistical machine translation. In *Association for Computational Linguistics*.

Forcada, M. L. and Ñeco, R. P. (1997). Recursive hetero-associative memories for translation. In J. Mira, R. Moreno-Díaz, and J. Cabestany, editors, *Biological and Artificial Computation: From Neuroscience to Technology*, volume 1240 of *Lecture Notes in Computer Science*, pages 453–462. Springer Berlin Heidelberg.

Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. (2013). Maxout networks. In *Proceedings of The 30th International Conference on Machine Learning*, pages 1319–1327.

Graves, A. (2012). Sequence transduction with recurrent neural networks. In *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv:*1308.0850 [cs.NE].

Graves, A., Jaitly, N., and Mohamed, A.-R. (2013). Hybrid speech recognition with deep bidirectional LSTM. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, pages 273–278.

ヘルマン、K.とブルンソム、P.（2014）.単語アライメントを用いない多言語分散表現。第2回学習表現国際会議(ICLR 2014)予稿集.

Hochreiter, S.（1991）. 動的な神経細胞ネットワークの研究。ディプロマ・テーゼ、情報学研究所、Lehrstuhl教授ミュンヘン工科大学、ブラウアー。

ホッホライター、S.とシュミッドフーバー、J.（1997）.長期短期記憶。神経計算，9(8)，1735-1780.

Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1700–1709. Association for Computational Linguistics.

Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press, New York, NY, USA.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 48–54, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pascanu, R., Mikolov, T., and Bengio, Y. (2013a). On the difficulty of training recurrent neural networks. In *ICML'2013*.

Pascanu, R., Mikolov, T., and Bengio, Y. (2013b). On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)*.

Pascanu, R., Gulcehre, C., Cho, K., and Bengio, Y. (2014). How to construct deep recurrent neural networks. In *Proceedings of the Second International Conference on Learning Representations (ICLR 2014)*.

Pouget-Abadie, J., Bahdanau, D., van Merriënboer, B., Cho, K., and Bengio, Y. (2014). Overcoming the curse of sentence length for neural machine translation using automatic segmentation. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. to appear.

Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, **45**(11), 2673–2681.

Schwenk, H. (2012). Continuous space translation models for phrase-based statistical machine translation. In M. Kay and C. Boitet, editors, *Proceedings of the 24th International Conference on Computational Linguistics (COLIN)*, pages 1071–1080. Indian Institute of Technology Bombay.

Schwenk, H., Dchelotte, D., and Gauvain, J.-L. (2006). Continuous space language models for statistical machine translation. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 723–730. Association for Computational Linguistics.

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems (NIPS 2014)*.

Zeiler, M. D. (2012). ADADELTA: An adaptive learning rate method. *arXiv:*1212.5701 [cs.LG].

モデル・アーキテクチャ

A.1 アーキテクチャの選択

セクション3で提案するスキームは、例えばリカレントニューラルネットワーク(RNN)の活性化関数fとアライメントモデルaを自由に定義できる一般的なフレームワークである。ここでは、本論文の実験に用いた選択について述べる。

A.1.1 現在のニューラルネットワーク

RNNの活性化関数fには、Choら(2014a)が最近提案したゲート付き隠れユニットを用いる。ゲーテッド隠れユニットは、要素ごとのtanhのような従来の単純なユニットに代わるものである。このゲート型ユニットは、Hochreiter and Schmidhuber (1997)によって以前に提案された長期短期記憶(LSTM)ユニットに似ており、長期依存関係をよりよくモデル化し学習する能力を共有している。これは、導関数の積が1に近い計算経路をアンフォールドRNNに持つことで可能となる。これらの経路は、消失効果にあまり悩まされることなく、勾配が容易に逆流することを可能にする(Hochreiter, 1991; Bengio et al., 1994; Pascanu et al., 2013a)。したがって、Sutskeverら(2014)が同様の文脈で行ったように、ここで説明したゲート付き隠れユニットの代わりにLSTMユニットを使用することが可能である。

n個のゲート付き隠れユニット[8]を採用したRNNの新しい状態s_iは次式で計算される。

$$s_i = f(s_{i-1}, y_{i-1}, c_i) = (1 - z_i) \circ s_{i-1} + z_i \circ \tilde{s}_i,$$

ここで、*は要素ごとの乗算、z_iは更新ゲートの出力である(下記参照)。提案する更新状態s_iは次式で計算される。

$$\tilde{s}_i = \tanh\left(We(y_{i-1}) + U\left[r_i \circ s_{i-1}\right] + Cc_i\right),$$

ここで、$e(y_{i-1}) \in R^m$ は単語 y_{i-1} の m 次元埋め込み、r_i はリセットゲートの出力である(下記参照)。y_i を1-of-Kベクトルとして表現する場合、$e(y_i)$ は単に埋め込み行列 $E \in R^{m \times K}$ の列である。可能な限り、バイアス項を省略し、方程式の乱雑さを少なくする。

更新ゲートz_iは各隠れユニットが前の活性化を維持することを可能にし、リセットゲートr_iは前の状態からの情報をどれだけ、どの情報にリセットすべきかを制御する。によって計算する。

$$z_i = \sigma\left(W_z e(y_{i-1}) + U_z s_{i-1} + C_z c_i\right),$$
$$r_i = \sigma\left(W_r e(y_{i-1}) + U_r s_{i-1} + C_r c_i\right),$$

ここで、$\sigma(\cdot)$ はロジスティックシグモイド関数である。

デコーダの各ステップで、出力確率(式(4))を多層関数として計算する(Pascanu et al., 2014)。maxout units (Goodfellow et al., 2013)の隠れ層を1層使用し、出力確率(各単語に1つずつ)をソフトマックス関数で正規化する(式(6)参照)。

A.1.2 整列モデル

アライメントモデルは、長さT_xとT_yの各文対に対して、$T_x \times T_y$回評価する必要があることを考慮して設計する必要がある。計算量を減らすために、次のような単層多層パーセプトロンを使用する。

$$a(s_{i-1}, h_j) = v_a^\top \tanh\left(W_a s_{i-1} + U_a h_j\right),$$

ここで、$W_a \in R^{n \times n}$、$U_a \in R^{n \times 2n}$、$v_a \in R^n$は重み行列である。$U_a\, h_j$ は i に依存しないので、計算コストを最小化するために事前に計算しておくことができる。

[8] ここでは、デコーダの式を示す。エンコーダでは、文脈ベクトルc iと関連する項を無視するだけで、同じ式を使用することができる。

A.2 D A.2 モデルの詳細説明

A.2.1 ENCODER

本節では、実験に用いた提案モデル(RNNsearch)のアーキテクチャを詳細に説明する(Sec.4-5参照)。以降、読みやすくするために、バイアス項をすべて省略する。

このモデルは、1-of-K符号化された単語ベクトルの原文を入力とする。

$$\mathbf{x} = (x_1, \ldots, x_{T_x}),\ x_i \in \mathbb{R}^{K_x}$$

となり、1-of-K符号化された単語ベクトルの翻訳文を出力する。

$$\mathbf{y} = (y_1, \ldots, y_{T_y}),\ y_i \in \mathbb{R}^{K_y},$$

ここで、K_xとK_yはそれぞれソース言語とターゲット言語の語彙サイズである。T_xとT_yはそれぞれ原文と訳文の長さを表す。

まず、双方向リカレントニューラルネットワーク(BiRNN)の順方向状態を計算する:

$$\overrightarrow{h}_i = \begin{cases} (1 - \overrightarrow{z}_i) \circ \overrightarrow{h}_{i-1} + \overrightarrow{z}_i \circ \overrightarrow{\underline{h}}_i & , \text{if } i > 0 \\ 0 & , \text{if } i = 0 \end{cases}$$

where

$$\begin{aligned} \overrightarrow{\underline{h}}_i &= \tanh\left(\overrightarrow{W}\overrightarrow{E}x_i + \overrightarrow{U}\left[\overrightarrow{r}_i \circ \overrightarrow{h}_{i-1}\right]\right) \\ \overrightarrow{z}_i &= \sigma\left(\overrightarrow{W}_z\overrightarrow{E}x_i + \overrightarrow{U}_z\overrightarrow{h}_{i-1}\right) \\ \overrightarrow{r}_i &= \sigma\left(\overrightarrow{W}_r\overrightarrow{E}x_i + \overrightarrow{U}_r\overrightarrow{h}_{i-1}\right). \end{aligned}$$

$\overrightarrow{E} \in \mathbb{R}^{m \times K}$ は単語埋め込み行列である。$\overrightarrow{W}$, $\overrightarrow{W}_z$, $\overrightarrow{W}_r \in \mathbb{R}^{n \times m}$, $\overrightarrow{U}$, $\overrightarrow{U}_z$, $\overrightarrow{U}_r \in \mathbb{R}^{n \times n}$ は重み行列である。mとnはそれぞれ単語埋め込み次元数と隠れユニット数である。$\sigma(\cdot)$は通常通りロジスティックシグモイド関数である。

$\overleftarrow{\ }$ 後方状態($\overleftarrow{h}_1$, $\overleftarrow{\ }$, $\overleftarrow{h}_{T_x}$)も同様に計算される。重み行列とは異なり、前方RNNと後方RNNの間で単語埋め込み行列Eを共有する。

前方状態と後方状態を連結して注釈($\overrightarrow{h}_1$, $\overrightarrow{h}_2$, $\overleftarrow{\ }$, $\overrightarrow{h}_{T_x}$)を得る。

$$h_i = \begin{bmatrix} \overrightarrow{h}_i \\ \overleftarrow{h}_i \end{bmatrix} \tag{7}$$

A.2.2 DECODER

エンコーダからのアノテーションが与えられたデコーダの隠れ状態s_iは次式で計算される。

$$s_i = (1 - z_i) \circ s_{i-1} + z_i \circ \tilde{s}_i,$$

where

$$\begin{aligned} \tilde{s}_i &= \tanh\left(WEy_{i-1} + U\left[r_i \circ s_{i-1}\right] + Cc_i\right) \\ z_i &= \sigma\left(W_zEy_{i-1} + U_zs_{i-1} + C_zc_i\right) \\ r_i &= \sigma\left(W_rEy_{i-1} + U_rs_{i-1} + C_rc_i\right) \end{aligned}$$

E はターゲット言語の単語埋め込み行列である。W, W_z, $W_r \in \mathbb{R}^{n \times m}$, U, U_z, $U_r \in \mathbb{R}^{n \times n}$, C, C_z, $C_r \in \mathbb{R}^{n \times 2n}$ は重みである。ここでも、mとnはそれぞれ単語埋め込み次元と隠れユニット数である。初期隠れ状態 s_0 は $s_0 = \overleftarrow{\ } \tanh W_s\ h_1$ で計算される。

文脈ベクトルc_iはアライメントモデルによって各ステップで再計算される:

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j,$$

Model	Updates ($\times 10^5$)	Epochs	Hours	GPU	Train NLL	Dev. NLL
RNNenc-30	8.46	6.4	109	TITAN BLACK	28.1	53.0
RNNenc-50	6.00	4.5	108	Quadro K-6000	44.0	43.6
RNNsearch-30	4.71	3.6	113	TITAN BLACK	26.7	47.2
RNNsearch-50	2.88	2.2	111	Quadro K-6000	40.7	38.1
RNNsearch-50*	6.67	5.0	252	Quadro K-6000	36.7	35.2

表2: 学習統計と関連情報。各更新は、1つのミニバッチを使用してパラメータを1回更新することに対応する。1エポックはトレーニングセットを1回通過することである。NLLは、訓練セットまたは開発セットのいずれかの文の平均条件付き対数確率である。文の長さが異なることに注意してください。

where

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$
$$e_{ij} = v_a^\top \tanh(W_a s_{i-1} + U_a h_j),$$

であり、h_jは原文のj番目のアノテーションである(式(7)参照)。$v_a \in \mathbb{R}^{n^0}$ 、$W_a \in \mathbb{R}^{n^0 \times n}$ 、$U_a \in \mathbb{R}^{n^0 \times 2n}$ は重み行列である。c_iをh_{Tx}に固定すると、モデルはRNNエンコーダ・デコーダ(Cho → – et al.

With the decoder state s_{i-1}, the context c_i and the last generated word y_{i-1}, we define the probability of a target word y_i as

$$p(y_i|s_i, y_{i-1}, c_i) \propto \exp\left(y_i^\top W_o t_i\right),$$

where

$$t_i = \left[\max\left\{\tilde{t}_{i,2j-1}, \tilde{t}_{i,2j}\right\}\right]_{j=1,\ldots,l}^\top$$

であり、$t \sim_{i,k}$ はベクトル $t \sim_i$ の k 番目の要素であり、次式で計算される。

$$\tilde{t}_i = U_o s_{i-1} + V_o E y_{i-1} + C_o c_i.$$

$W_o \in \mathbb{R}^{K\ y \times l}$, $U_o \in \mathbb{R}^{2l \times n}$, $V_o \in \mathbb{R}^{2l \times m}$, $C_o \in \mathbb{R}^{2l \times 2n}$ は重み行列である。これは、単一のマックスアウト隠れ層(Goodfellow et al., 2013)を持つ深い出力(Pascanu et al., 2014)を持つと理解できる。

A.2.3 モデルサイズ

本論文で使用する全てのモデルにおいて、隠れ層のサイズnは1000、単語埋め込み次元mは620、深層出力lの最大出力隠れ層のサイズは500である。アライメントモデルn^0の隠れユニット数は1000である。

B トレーニング手順

B.1 パラメータの初期化

← – ← – ← – → – → – リカレント重み行列 U, U_z, U_r, U, U_z, U_r, U, U_z, U_r をランダムな直交行列として初期化した。W_a と U_a については、平均 0、分散 0.001^2 のガウス分布から各要素をサンプリングして初期化した。V_a のすべての要素とすべてのバイアス・ベクトルはゼロに初期化された。その他の重み行列は、平均0、分散0.01^2のガウス分布からサンプリングして初期化した。

B.2 TRAINING

確率的勾配降下(SGD)アルゴリズムを使用した。Adadelta (Zeiler, 2012)を用いて、各パラメータの学習率を自動的に適応させた(= 10^{-6} , ρ = 0.95)。

我々は、ノルムが閾値より大きいとき、コスト関数の勾配のL_2 -ノルムを毎回明示的に正規化し、最大でも予め定義された閾値1になるようにした(Pascanu et al., 2013b)。各SGD更新方向は、80文のミニバッチで計算された。

各更新において、我々の実装はミニバッチ内の最長文の長さに比例した時間を必要とする。したがって、計算の無駄を最小限に抑えるため、20回目の更新の前に、1600文のペアを検索し、長さに従ってソートし、20のミニバッチに分割した。学習データは学習前に1回シャッフルされ、このように順次トラバースされた。

表2に、実験に使用したすべてのモデルのトレーニングに関する統計量を示す。

ロング・エントランスの翻訳

Source	An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.
Reference	入院の特権は、病院を収容する会員の地位を通じた医師が、患者を病院または医療センターに入院させ、診断または治療を行わせる権利である。
RNNenc-50	Un privilège d'admission est le droit d'un médecin de reconnaître un patient à l'hôpital ou un centre médical d'un diagnostic ou de prendre un diagnostic en fonction de son état de santé.
RNNsearch-50	Un privilège d'admission est le droit d'un médecin d'admettre un patient à un hôpital ou un centre médical pour effectuer un diagnostic ou une procédure, selon son statut de travailleur des soins de santé à l'hôpital.
Google Translate	入院の特権とは、病院または医療センターで患者を受け入れ、病院における健康管理に努める限り、その状況に基づいて診断または処置を行う権利である。

Source	このような経験は、ディズニーが「シリーズの寿命を延ばし、デジタル・プラットフォームを通じてオーディエンスと新たな関係を構築し、ますます重要性を増している」という取り組みの一環である。
Reference	この種の経験は、ディズニーの努力の一環として、「シリーズの存続期間を延長し、ますます重要なデジタルプラットフォームのおかげで、彼の公衆と新たな関係を構築する」ために行われた、と彼は付け加えた。
RNNenc-50	Ce type d'expérience fait partie des initiatives du Disney pour "prolonger la durée de vie de ses nouvelles et de développer des liens avec les lecteurs numériques qui deviennent plus complexes.
RNNsearch-50	この種の経験は、ディズニーが「シリーズの存続期間を延長し、より重要なデジタルプラットフォームを通じて、一般の人々との新しい関係を構築する」という努力の一環であった。
Google Translate	この種の経験は、ディズニーが「そのシリーズの存続期間を延長し、ますます重要になるデジタル・プラットフォームの恩恵によって、一般の人々との新しい関係を構築する」ことに努めたことの一部である、と彼は付け加えた。

Source	In a press conference on Thursday, Mr Blair stated that there was nothing in this video that might constitute a "reasonable motive" that could lead to criminal charges being brought against the mayor.
Reference	En conférence de presse, jeudi, M. Blair a affirmé qu'il n'y avait rien dans cette vidéo qui puisse constituer des "motifs raisonnables" pouvant mener au dépôt d'une accusation criminelle contre le maire.
RNNenc-50	jeudiの報道会議で、M.ブレアは、市長に対する犯罪告発に巻き込まれる可能性のある「合理的動機」を構成する可能性のあるこのビデオに、何もなかったと述べた。
RNNsearch-50	ブリード・ミーティングで、M.ブレアは、市長に対する罪の非難につながる「理性的な動機」を構成する可能性のあるこの流行に、何もなかったと宣言した。
Google Translate	ブリード・ミーティングで、M.ブレアは、市長に対する犯罪告発に言及できる「理性的な動機」を構成する可能性のあるこの生活には何もないと宣言した。

表3:テストセットから選択された長い原文(30語以上)からRNNenc-50とRNNsearch-50が生成した翻訳。各原文について、ゴールドスタンダード翻訳も示す。Google翻訳による翻訳は2014年8月27日に行われた。