

リカレントネットワークを用いた機械翻訳の試み

An attempt to apply a recurrent network to machine translation

住井 泰介*1 岡 夏樹*2
Taisuke SUMII Natsuki OKA

*1 京都工芸繊維大学 工学学部 電子情報工学科

Department of Electronics and Information Science, Faculty of Engineering and Design, Kyoto Institute of Technology

*2 京都工芸繊維大学 大学院 工学科学研究科 情報工学部門

Department of Information Science, Graduate School of Science and Technology, Kyoto Institute of Technology

We applied a recurrent network (SRN) to machine translation. The network learns to translate some simple English sentences to Japanese from a set of examples of translation. We demonstrate that the network has ability to translate sentences which include unknown combinations of words even if the word order or the number of words differs between the source sentence and the target sentence.

1. はじめに

計算機に“意味”を扱わせようとするのは、歴史ある人工知能研究における根本的問題の一つである。そこへのアプローチの一つに、例えば記号論理を基にしたもの [1] があり、あるいは、より脳の生物学的構造に忠実であろうとしたものがある。後者に属するものとしてニューラルネットワークを用いたアプローチがあり [2]、そのうちの一つに、Elman による、リカレントネットワークによる言語処理に関する研究がある [3][4][5]。

本研究では、このリカレントネットワークを用いた言語処理研究手法を翻訳というタスクに適用することを試み、リカレントネットワークによる言語処理の応用可能性を探求するとともに、計算機による言語理解への道筋の探求の一助とすることを目的とする。

2. 機械翻訳とリカレントネット

2.1 既存の機械翻訳

機械翻訳の実現は、長い歴史を持つ人工知能研究のドライビング・フォースの一つであり続けてきた重要なテーマである。そこでは、初期においては、文法などの言語的ルールを自然言語文の解析あるいは生成を行う機構に明示的に組み込む手法が用いられたが、現在では統計的手法を用いるものが主流となっている [6]。

2.2 リカレントネットとそれを用いた言語関連研究

リカレントネットはニューラルネットの一形態であり、階層型でありながら、後段から前段への何らかのフィードバック構造を備えたものである。それにより、ニューラルネットのパターン認識機能を系列信号に適用することができ、また、順次、出力とフィードバックを繰り返すことにより、系列再生能力も実現している。

リカレントネットを用いた言語関連研究には、言語獲得や表象といった事柄を、ユニット値の状態空間上の遷移をたどることによって考察しようとしたもの [7] や、学習内容・過程を解析することにより考察しようとしたもの [3] などがある。

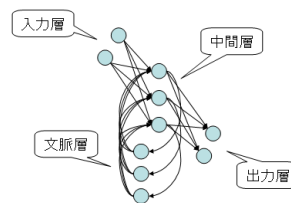


図 1: 本研究で用いたリカレントネットワーク。エルマンの単純回帰ネットワーク (SRN: Simple Recurrent Network) と同等。

2.3 本研究の位置づけ

本研究では機械翻訳にリカレントネットを適用するものの、直ちに既存の機械翻訳の翻訳水準に達するものを期待するものではなく、あくまで翻訳に利用できる“可能性”を探る段階のものである。むしろ第一の目的は、このようなタスクを実行する際にリカレントネットがどのような振る舞いをとることにより問題を解決しているかを探り、計算機による言語理解への道を探ることにある。

3. 本研究で用いた翻訳用リカレントネット

3.1 用いたリカレントネット

用いたリカレントネット (SRN: Simple Recurrent Network) の模式図を図 1 に示す。

3.2 学習データの与え方

入力信号と教師信号の与え方を図 2 に示す。

以下、学習データセット中の、ある一つの被翻訳文・翻訳文ペアを学習させる方法を述べる。

まず、被翻訳文の一単語を入力信号ベクトル及び教師信号ベクトルとして同時に用い、バックプロパゲーションする。この際、バックプロパゲーションは、RTRL 法にならぬ、一つの教師信号ベクトルが与えられるたびに、その時点での入力層及び一時点前の中間層を終着点として遡及する。さらに、入力信号ベクトル・教師信号ベクトルに順次次の単語を用い、これを被翻訳文中の単語が続く限り進める。

被翻訳文の単語を用い終えたら、続けて、入力信号は無しとして、翻訳文の単語を教師信号にして、同じくバックプロパ

連絡先: 住井 泰介, 京都工芸繊維大学 工学学部 電子情報工学科,
京都市左京区松ヶ崎御所海道町, m7622023@edu.kit.ac.jp

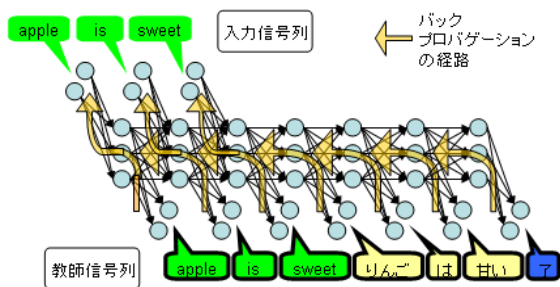


図 2: 学習データの与え方。文脈層の状態は前時点での中間層の状態そのままなので省略。一単語につき一入力ベクトル・教師信号ベクトルにしたのち、入力信号なし、かつ教師信号は翻訳文の各単語、とする。その後、翻訳終了を示す単語（図中の「了」）を教師信号として与えて終了。学習は RTRL 法にならう。

ゲーションする。これを翻訳文の単語を一通り教師信号として使い切るまで続ける。

最後に、翻訳終了を示す単語を教師信号に使い、この被翻訳文・翻訳文ペアによる一回の学習は終了する。

4. 翻訳実験

4.1 翻訳実験の目的

この実験の目的は、リカレントネットを用いての機械翻訳がどの程度可能であるかを把握することである。また同時に、その際のより良い学習データとはいかなるものか、を探ることも行う。

4.2 概略

英語文を日本語文に翻訳させる実験である。実際の翻訳に先立ち、被翻訳文・翻訳文ペアなどからなる学習データセットを用いて学習を行う。学習後の翻訳実験には、当然ながら、学習データセットには含まれていない文章を被翻訳文に用いる。

なお、リカレントネットの入力・中間・出力の各層のユニット数は、すべて各実験で用いる総単語数（被翻訳文と翻訳文ともに含む）と同数である。入力層と出力層では、一ユニットが一単語に対応している。また、学習係数は 0.1 とした。

4.3 実験 1: 文法的に比較的単純な場合

ここでは、表 1 などに示すデータを用いて学習後、“apple is bitter” を翻訳させ、“りんご は 甘い 了” という翻訳結果を期待する。ここで、“了” は翻訳終了を示す単語である。学習回数が 1000 回加わるとに翻訳実験をし、それを学習回数が 20000 回になるまで繰り返した。その通し実験を各小実験で計 20 回行った。

実験 1.1: 完全な文のみを学習に用いる場合。学習データセットを表 1 に示す。結果は表 2 に示すように、全 400 回の翻訳実験（20 試行×各 20 実験）において、期待する翻訳結果が得られたのが 22 回。

実験 1.2: 完全な文に加え、文節単位や単語単位の学習データも用いる場合。学習データセットを表 3 に示す。結果は、全 400 回の翻訳実験において、期待する翻訳結果が得られたのが 27 回。これは実験 1.1 の結果より増大している。

実験 1.3: 学習に完全な文のみを用いながら、主語となる単語が複数の述語を持ちうる場合。学習データセットを表 4 に

学習用被翻訳文	対応する学習用翻訳文
apple is sweet	りんご は 甘い 了
medicine is bitter	薬 は 苦い 了

表 1: 実験 1.1 の学習データセット。実験 1 はすべて比較的単純な翻訳で、学習後、“apple is bitter” を翻訳させる。実験 1.1 はその上でさらに、学習データに完全な文のみを用いる場合。翻訳文の日本語は、空白で区切られているのが単語。一単語にリカレントネットワークへの一入力信号ベクトル・教師信号ベクトルを充てる。被翻訳文の単語は入力信号にも教師信号にも用いる。翻訳文の単語は教師信号にのみ用いる。なお、「了」は翻訳終了を示す単語。

試行	翻訳結果（学習が 1000 回経過ごとに翻訳）
1	××××××××××××××××××××××××××××
2	××××××××××××××××××××××××××××
3	× ××××××××××××××××××××××××××××
4	×××××××
5	××××××××××××××××××××××××××××
6	××××××××××××××××××××××××××××
7	× ××××××××××××××××××××××××××××
8	××××××××××××××××××××××××××××
9	××××××××××××××××××××××××××××
10	× ××××××××××××××××××××××××××××
11	××××××××××××××××××××××××××××
12	××××××××××××××××××××××××××××
13	××××××××××××××××××××××××××××
14	××××××××××××××××××××××××××××
15	××××××××××××××××××××××××××××
16	××××××××××××××××××××××××××××
17	×××××××× ××××××××××××××××××××
18	××××××××××××××××××××××××××××
19	××××××××××××××××××××××××××××
20	××××××××××××××××××××××××××××

表 2: 実験 1.1 の結果。望む翻訳結果は“りんご は 苦い 了”。左から順に、学習開始以来、学習回数が 1000 回加わるとの結果を示す。は望んだ翻訳結果が得られたこと、×は得られなかったことを示す。表中、は総数 22。

学習用被翻訳文	対応する学習用翻訳文
apple is sweet	りんご は 甘い 了
medicine is bitter	薬 は 苦い 了
apple is	りんご は 了
medicine is	薬 は 了
is sweet	は 甘い 了
is bitter	は 苦い 了
apple	りんご 了
medicine	薬 了
sweet	甘い 了
bitter	苦い 了

表 3: 実験 1.2 の学習データセット。完全な文に加え、文節単位や単語単位のものも学習データに用いる場合。実験 1.1 に比べて翻訳成功率が向上する（本文参照）。

学習用被翻訳文	対応する学習用翻訳文
apple is sweet	りんご は 甘い 了
medicine is bitter	薬 は 苦い 了
choco is sweet	チョコ は 甘い 了
choco is bitter	チョコ は 苦い 了

表 4: 実験 1.3 の学習データセット。学習に完全な文のみを用いながらも、主語となる単語が複数の文中で用いられ、複数の述語を持つことがある場合。実験 1.1 に比べて翻訳成功率が向上する (本文参照)。

学習用被翻訳文	対応する学習用翻訳文
apple is sweet	りんご は 甘い 了
medicine is bitter	薬 は 苦い 了
chocolate is sweet	チョコ は 甘い 了
chocolate is bitter	チョコ は 苦い 了
apple is	りんご は 了
medicine is	薬 は 了
chocolate is	チョコ は 了
is sweet	は 甘い 了
is bitter	は 苦い 了
apple	りんご 了
medicine	薬 了
chocolate	チョコ 了

表 5: 実験 1.4 の学習データセット。学習データとして完全な文に加えて文節単位や単語単位のものも用い、さらに、一つの主語単語が複数の述語をもつことのある場合。実験 1.2、1.3 両方に比べて翻訳成功率が向上する (本文参照)。

示す。結果は、全 400 回の翻訳実験において、期待する翻訳結果が得られたのが 38 回。これは実験 1.1 の結果を上回る。

実験 1.4: 学習に完全な文に加えて文節単位や単語単位のものも用い、さらに一つの主語単語が複数の述語を持ちうる場合。学習データセットを表 5 に示す。結果は全 400 回の翻訳実験において、期待する翻訳結果が得られたのが 48 回。これは実験 1.2 及び 1.3 の結果をとともに上回る。

考察: 文法的に比較的単純な翻訳タスクにおいて、翻訳がなすことが示された。また、翻訳に成功した回の分布は累計学習回数の大小との相関が比較的低いため (実験 1.1 以外の結果表は割愛) 各小実験の比較より、学習データは文のみからなるより文節単位や単語単位を含むほうが、また、交差的に用いられている単語が含まれているほうが、効果的な学習データとなることが示された。

4.4 実験 2: 被翻訳文・翻訳文間で語数・語順が異なる場合

ここでは、表 6 に示すデータを用いて学習後、“I go to home” を翻訳させ、“私は家へ行く了” という翻訳結果を期待する。学習回数が 1000 回加わることにより翻訳実験をし、それを学習回数が 20000 回になるまで繰り返した。その通り実験を計 20 回行った。なお、これ以降の実験では、学習データには文節単位や単語単位のものも含め、かつ、交差的に用いる単語を含める。

結果: 全 400 回の翻訳実験において、期待する翻訳結果が得られたのが 26 回。

考察: 語順と語数が異なる場合においても翻訳がなすことが

学習用被翻訳文	対応する学習用翻訳文
I go to school	私 は 学校 へ 行く 了
you go to home	あなた は 家 へ 行く 了
he go to school	彼 は 学校 へ 行く 了
he go to home	彼 は 家 へ 行く 了
I	私 了
you	あなた 了
he	彼 了
to school	学校 へ 了
to home	家 へ 了
go to school	学校 へ 行く 了
go to home	家 へ 行く 了

表 6: 実験 2.1 の学習データセット。被翻訳文・翻訳文間で語数・語順が異なる場合。学習後、“I go to home” を翻訳させる。

とが示された。

4.5 実験 3: 被翻訳文の語順の変化に翻訳文を対応させて変化させる必要がある場合

ここでは、表 7 など示すデータを用いて学習後、“he is teacher” 及び “is he teacher” あるいは “he is teacher” 及び “is he teacher?” を翻訳させ、“彼は教師です” 及び “彼は教師ですか” という翻訳結果を期待する。学習回数が 1000 回加わることにより翻訳実験をし、それを学習回数が 20000 回になるまで繰り返した。その通り実験を各小実験で計 10 回行った。

実験 3.1: 単語として “?” を用いない場合。学習データセットを表 7 に示す。結果は、全 200 回の翻訳実験において、2 つの被翻訳文のうち片方でも期待する翻訳結果が得られたのが 34 回。両方同時に得られたのが 8 回。なお、“he is teacher” のみ数えると 20 回。“is he teacher” なら 22 回。

実験 3.2: 単語として “?” を用いる場合。学習データセットを表 8 に示す。結果は、全 200 回の翻訳実験において、2 つの被翻訳文のうち片方でも期待する翻訳結果が得られたのが 97 回。両方同時に得られたのが 34 回。これらの数はともに実験 3.1 に比べて増大している。

考察: 実験 3.1 より、被翻訳文間に語順の違いしかない場合にも、それに対応した翻訳をなすことが示された。また、被翻訳文と翻訳文の間で語順が異なる場合と異なる場合では、期待する翻訳の達成率にはさほどの違いがないことも示された (実験 3.1 の結果を参照)。また、実験 3.1 と 3.2 の結果の対比より、被翻訳文間に語順の違いしかない場合には、文書の類型を際立たせる単語 (ここでは “?”) を導入すると、期待する翻訳の達成率が格段に向上することが示された。

4.6 実験全体の考察

以上の実験により、

- 学習データのあり方、用い方により、学習後の翻訳達成率に明らかな差が生じること
- 文法的に極めて単純な翻訳以外の翻訳にも対応しうること
- 従って、文法的構造を複雑化・高度化しても、学習データを工夫すればそれに対応した翻訳ができる可能性があること

が示された。

学習用被翻訳文	対応する学習用翻訳文
she is teacher	彼女は教師です
is she teacher	彼女は教師ですか
he is student	彼は生徒です
is he student	彼は生徒ですか
ted is teacher	テッドは教師です
is ted teacher	テッドは教師ですか
ted is student	テッドは生徒です
is ted student	テッドは生徒ですか
bob is teacher	ボブは教師です
is bob teacher	ボブは教師ですか
bob is student	ボブは生徒です
is bob student	ボブは生徒ですか
she	彼女
he	彼
ted	テッド
bob	ボブ
teacher	教師
student	生徒

表 7: 実験 3.1 の学習データセット。被翻訳文間の語順の違いに対応する必要があり、さらに、“?” という単語を用いない場合。学習後、“he is teacher” と “is he teacher” を翻訳させる。

学習用被翻訳文	対応する学習用翻訳文
she is teacher	彼女は教師です
is she teacher ?	彼女は教師ですか
he is student	彼は生徒です
is he student ?	彼は生徒ですか
ted is teacher	テッドは教師です
is ted teacher ?	テッドは教師ですか
ted is student	テッドは生徒です
is ted student ?	テッドは生徒ですか
bob is teacher	ボブは教師です
is bob teacher ?	ボブは教師ですか
bob is student	ボブは生徒です
is bob student ?	ボブは生徒ですか
she	彼女
he	彼
ted	テッド
bob	ボブ
teacher	教師
student	生徒

表 8: 実験 3.2 の学習データセット。被翻訳文間の語順の違いに対応する必要があり、さらに、“?” という単語も用いる場合。学習後、“he is teacher” と “is he teacher ?” を翻訳させる。実験 3.1 に比べて翻訳成功率が向上する (本文参照)。

5. 意味ネットワークとしてのリカレントネットについての考察

意味ネットワークは、一般的には離散的グラフの形態をとる [8]。概念グラフも同様である。

一方、人間の認知作用においては、離散的には割り切れない概念や思考がその大きな部分を占めており、論理的命題でさえも、連続値的曖昧さを内包するものだと我々は考える。また、意味や概念といったものの主な作用は、その内容の展開ないし推移であるとも考える。

ここで、リカレントネットは、連続値によりその特性が構成されるものであり、かつ、言語処理に用いられる際には、言語系列的な展開ないし推移を実現する。

従って、リカレントネットは意味や概念を扱うための機構として利用可能であり、また、本研究によってさらに広い言語処理機能が確かめられたことにより、その可能性もより強く確かめられたものとする。

6. おわりに

本試みでは、極めて単純なタスクかつ小規模な学習データしか扱えなかった。

今後は、大規模なデータでの学習の後、比較的高度な翻訳タスクを試行したい。

参考文献

- [1] Michael R. Genesereth, Nils J. Nilsson: Logical Foundations of Artificial Intelligence, Morgan Kaufmann Publishers (1987)
- [2] D.E. Rumelhart, J.L. McClelland and the PDP Research Group: Parallel distributed processing: explorations in the microstructure of cognition, Volume 1, 2, The MIT Press. (1986)
- [3] Elman, J.L.: Finding structure in time, Cognitive Science, 14, pp179 - 211. (1991)
- [4] Elman, J.L.: Distributed representations, simple recurrent networks, and grammatical structure, Machine Learning, 7, pp195-224. (1991)
- [5] Elman, J.L.: Learning and development in neural networks: The importance of starting small, Cognition, 48, pp71 - 99. (1991)
- [6] 長尾 真: 第 1 章自然言語処理の歴史 自然言語処理 岩波講座ソフトウェア科学 15, pp1 - 12, 岩波書店, 第 7 刷. (2003)
- [7] Servan-Schreiber, D., Cleeremans, A. McClelland, J.L.: Graded State Machines: the representation of temporal contingencies in Simple Recurrent Networks, Machine Learning, 7, pp161-193. (1989)
- [8] G. A. Ringland and D. A. Duce: Knowledge Representation - An Introduction, Research Studies Press LTD. (1988)