G-021

リカレントネットを用いた翻訳と記号操作的処理の探求

Translation and Symbol Manipulation with a Reccurrent Network

住井 泰介 †　　**岡 夏樹** ‡
Taisuke SUMII　Natsuki OKA

1　はじめに

本研究は、ニューラルネットの一形態であるリカレントネットを用いての機械翻訳及び記号変換を試行し、“計算機による言語理解” への道筋を探求する一助とすることを目的とする。

計算機に “言葉の意味” を扱わせようとすることは、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に、記号論理を基にしたもの [1] や、より脳の生物学的構造に忠実であろうとしたもの、つまり、ニューラルネットを用いたアプローチがあり [2]、さらにそのうちの一つに Elman によるリカレントネットによる言語処理に関する研究がある [3][4][5]。

本研究の特質は、このリカレントネットを学習の後、“未学習の系列信号（被翻訳文）” を入力し、パターン認識処理を経て、“期待する系列信号出力（翻訳文）” が得られるかどうかを実験することにある。

2　機械翻訳とリカレントネット

2.1　既存の機械翻訳

翻訳という作業は知識や言語の統合処理に基づくものである。そのため、機械翻訳は人工知能研究における総合技術的な分野として扱われてきた。

そこでは、初期においては文法などの言語的ルールを文の解析や生成を行う機構に明示的に組み込む手法が用いられていたが、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　リカレントネットを用いた言語関連研究

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り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って考察しようとしたもの [3] などがある。

2.3　本研究の位置づけ

本研究の究極目的は、計算機による言語理解である。従って、機械翻訳タスクは、その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かを探ることが、その具体的な目的である。従って、直ちに既存の機械翻訳の翻訳水準を達成することを期待するものではない。

† 京都工芸繊維大学 大学院工芸科学研究科 情報工学専攻, Division of Information Science, Graduate School of Science and Technology, Kyoto Institute of Technology; m7622023@edu.kit.ac.jp
‡ 京都工芸繊維大学 大学院工芸科学研究科 情報工学部門, Departmrnt of Information Science, Graduate School of Science and Technology, Kyoto Institute of Technology

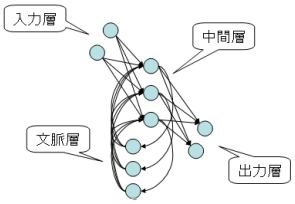

図 1　本研究で用いたリカレントネット。Elman の単純回帰ネットワーク（SRN: Simple Reccurent Network）と同等。

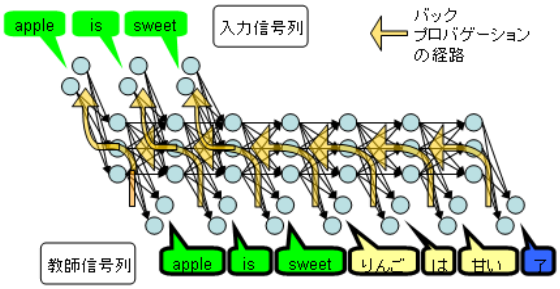

図 2　学習データの与え方。文脈層の状態は前時点での中間層の状態そのままなので省略。一単語につき一入力ベクトル・教師信号ベクトルを充てる。まず被翻訳文の各単語を入力信号及び教師信号にしたのち、入力信号なし、かつ教師信号は翻訳文の各単語、とする。その後、翻訳終了を示す単語（図中の「了」）を教師信号として与えて終了。学習は RTRL 法にならう。なお、教師信号として被翻訳文の単語も用いるのは、そうでない限り、被翻訳文の単語列パターンに対するリカレントネットの学習が十分でないと推測した（未確認）ため。

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の模式図を図 1 に示す。

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。

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、RTLT 法にならう）。さらに、入力信号ベクトル・教師信号ベク

表 1　実験 A1.1 の学習データセット。実験 A1 では学習後、“apple is bitter” を翻訳させる。なかでも実験 A1.1 は比較的少量の学習データを用いる場合。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。

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

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。

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用い、この被翻訳文・翻訳文ペアによる一回の学習は終了する。

なお、教師信号として翻訳文の単語のみならず被翻訳文の単語も用いるのは、そうでない限り、被翻訳文の単語列パターンに対するリカレントネットの学習が十分でないと推測した（未確認）からである。

4　英日翻訳実験（実験 A）

4.1　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。

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。

4.3　実験 A1：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　苦い　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 回加わるごとに翻訳実験をし、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 100 回行った。なお、各層のユニット数は全て、A1.1、1.2 共に 19。

実験 A1.1：比較的少量のデータ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 2000 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（平均 0.38 回/20 回中、標準偏差 1.76 回）。

実験 A1.2 : 比較的多量の学習データを用いる場合。学習データセットを表 3 に示す。結果は、全 2000 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 209 回（平均 2.1 回/20 回中、標準偏差 3.3 回）。これは実験 A1.1 の結果を有意に上回る。

考察：比較的単純な翻訳タスクにおいて、翻訳がなしうることが示された。また、翻訳実験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（表 2 の各試行下段（「実験」）を参照。実験 A1.1 以外の結果表は割愛）、実験 A1.1 と A1.2 の比較より、学習データには多くの例文を用いた方が、翻訳成功率が向上することが明らかになった。

表 2　実験 A1.1 の結果の一部。実際の全試行数は 100。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 回加わるごとの結果を示す。「学習」は学習データを全て学習できているか、「実験」は実験用データによる実験結果。○は望んだ結果が得られたこと、×は得られなかったことを示す。

		翻訳結果（学習が 1000 回経過ごとに翻訳）
1	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
2	学習	×○×○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
3	学習	○○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
4	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	××××××○○○○○○○○○○○○○○
5	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
6	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
7	学習	×○○○×○○○○○○○○○○○○○○○○
	実験	×○×××××××××××××××××××
8	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
9	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×××××××××××××××××××××
10	学習	×○○○○○○○○○○○○○○○○○○○○
	実験	×○○×××××××××××××××××

表 3　実験 A1.2 の学習データセット。比較的多量の学習データを用いる場合。実験 A1.1 に比べて翻訳成功率が向上する（本文 4.3 節参照）。

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
apple is red	りんご　は　赤い　了
medicine is white	薬　は　白い　了
cake is sweet	ケーキ　は　甘い　了
cocoa is bitter	ココア　は　苦い　了

4.4　実験 A2：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 4 などに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 了” 及び “彼 は 教師 です か 了” という翻訳結果を期待する。学習回数が 1000 回加わるごとに翻訳実験をし、それを学習回数が 20000 回になるまで繰り返した。その通し

表 4　実験 A2.1 の学習データセット。実験 A2 は被翻訳文間の語順の違いに対応する必要がある場合。なかでも A2.1 は、“?” 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 了
is she teacher	彼女 は 教師 です か 了
he is student	彼 は 生徒 です 了
is he student	彼 は 生徒 です か 了
ted is teacher	テッド は 教師 です 了
is ted teacher	テッド は 教師 です か 了
ted is student	テッド は 生徒 です 了
is ted student	テッド は 生徒 です か 了
bob is teacher	ボブ は 教師 です 了
is bob teacher	ボブ は 教師 です か 了
bob is student	ボブ は 生徒 です 了
is bob student	ボブ は 生徒 です か 了
she	彼女 了
he	彼 了
ted	テッド 了
bob	ボブ 了
teacher	教師 了
student	生徒 了

表 5　実験 A2.2 の学習データセット。“?” 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 A2.1 に比べて翻訳成功率が向上する（本文 4.4 節参照）。

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 了
is she teacher ?	彼女 は 教師 です か 了
he is student	彼 は 生徒 です 了
is he student ?	彼 は 生徒 です か 了
ted is teacher	テッド は 教師 です 了
is ted teacher ?	テッド は 教師 です か 了
ted is student	テッド は 生徒 です 了
is ted student ?	テッド は 生徒 です か 了
bob is teacher	ボブ は 教師 です 了
is bob teacher ?	ボブ は 教師 です か 了
bob is student	ボブ は 生徒 です 了
is bob student ?	ボブ は 生徒 です か 了
she	彼女 了
he	彼 了
ted	テッド 了
bob	ボブ 了
teacher	教師 了
student	生徒 了

実験を各小実験で計 10 回行った。

実験 A2.1：単語として “?” を用いない場合。学習データセットを表 4 に示す。結果は、全 200 回の翻訳実験において、2 つの被翻訳文から同時に期待する翻訳結果が得られたのが 8 回（平均 0.8 回/20 回中、標準偏差 1.4 回）。なお、“he is teacher” のみ数えると 20 回。“is he teacher” なら 22 回。

実験 A2.2：単語として “?” を用いる場合。学習データセットを表 5 に示す。結果は、全 200 回の翻訳実験において、2 つの被翻訳文から同時に期待する結果が得られたのが 34 回（平均 3.6 回/20 回中、標準偏差 3.5 回）。これらの数はともに実験 A2.1 の結果を上回る。

考察：実験 A2.1 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しうることが示された。また、被翻訳文と翻訳文の間で語順が異なる場合と異ならない場合では、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 A2.1 の結果を参照）。また、実験 A2.1 と 2.2 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。

4.5 実験 A 全体の考察

以上の実験により、リカレントネットによる機械翻訳に関し、

- 文法的に極めて単純な翻訳をなしうること
- 学習データの量を増大させ、或いは、質に工夫を加えると翻訳達成率が向上すること
- 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ると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。

5 記号操作的翻訳実験（実験 B）

5.1 目的

この実験の目的は、被翻訳文と翻訳文の対応関係に、より記号操作的側面を強調した学習・実験データを用い、その記号操作的処理能力を評価することである。

ニューラルネットを用いて記号操作的処理をすることは、過去に多くの研究がなされたが、なお十分な成功は達成されていない。ここでは、本研究の特質である、未学習の系列入力に対し望んだ系列出力を期待する、という方法のもと、この問題に取り組む。

5.2 実験

この実験では表 6 に示すデータを用いて学習後、“c a b” を翻訳させ “C A B” という翻訳結果を期待する。入力層、出力層のユニット数はともに 6。中間層ユニット数は 50。学習回数が 100 回加わるごとに翻訳実験をし、それを学習回数が 10000 回になるまで繰り返した。その通し実験を計 20 回行った。結果は、全 2000 回の翻訳実験において、期待する結果が得られたのが 548 回（平均 27.4 回/100 回中、標準偏差 15.6 回）。

5.3 考察

3 種の記号を 3 個並べる組み合わせは 27 通りであり、(1/27) × 100=3.7 であるが、実験では平均して 100 回中 27.4 回成功している。これはリカレントネットが記号操作的処理能力を有している可能性を示している。

6 おわりに

本研究では、極めて単純なタスクかつ小規模なデータしか扱えなかった。

表 6　実験 B の学習データセット。記号操作的側面が強い場合。アルファベット一文字が実験 A での一単語に相当。“c a b”、“C A B” ペアが含まれていないことに注意。また、タスクの簡単化のため “了” を用いていないことにも注意。学習後、“c a b” を翻訳させる。

学習用被翻訳文	対応する学習用翻訳文
a	A
b	B
c	C
a a	A A
a b	A B
a c	A C
b a	B A
b b	B B
b c	B C
c a	C A
c b	C B
c c	C C
a a a	A A A
a a b	A A B
a a c	A A C
a b a	A B A
a b b	A B B
a b c	A B C
a c a	A C A
a c b	A C B
a c c	A C C
b a a	B A A
b a b	B A B
b a c	B A C
b b a	B B A
b b b	B B B
b b c	B B C
b c a	B C A
b c b	B C B
b c c	B C C
c a a	C A A
c a c	C A C
c b a	C B A
c b b	C B B
c b c	C B C
c c a	C C A
c c b	C C B
c c c	C C C

今後は大規模なデータでの学習の後、比較的高度な自然言語文翻訳タスクを試行するとともに、応用的な記号操作的処理能力の探求を行いたい。また、内部状態の数理的解析、及び、数理的モデルの構築にも取り組みたい。

参考文献

[1] Michael R. Genesereth, Nils J. Nislsson: 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 countingencies in Simple Reccurrent Networks, Machine Learning, 7, pp161-193. (1989)