

階層型ニューラルネットによる深層格解析

本 木 実[†] 嶋 津 好 生[†] 高 橋 直 人^{††}

本論文では、階層型ニューラルネットワークを用いた新しい深層格解析の手法を提案する。高橋らの提案したネットワーク CAN は文の深層格解析を行いながら、単語に対応した実数ベクトル(単語表現ベクトル)を獲得することが可能である。我々は、この CAN を基に文が意味的に適格か否かを判定するニューラルネットを実現した。名詞 24 語、動詞 4 語を用い、計算機で構文的に正しい文を 57701 文作成し、そのうち意味的にも正しいと受け入れられる文 1275 文を選択制限性を満たす文として定義することで選択制限性の判定実験を行った。クロスバリデーションテストによって、学習用データに対する学習能力および未学習データに対する一般化能力を調べたところ、正答率はそれぞれ 99.59%、95.99%であった。また、学習用データの量が一般化能力に影響を及ぼすことが確かめられた。

Deep Case Analysis Using a Layered Neural Network

MINORU MOTOKI,[†] YOSHIO SHIMAZU[†] and NAOTO TAKAHASHI^{††}

This paper proposes a new method for case structure analysis using a neural network. The neural network model CAN, presented by Takahashi et al., performs case-role assignment while acquiring real-valued vectors (word representation vectors) corresponding to words. We developed CAN into a new neural network that judges selectional restriction of a sentence. We tested the performance of our neural network with 57701 sentences; these sentences were automatically generated from 24 nouns and 4 verbs, and 1275 sentences out of them were defined as semantically acceptable. The percentages of correct answers were 99.59% for training data and 95.99% for test data. Moreover, it was confirmed that the number of sentences used in learning affects generalization ability of analysis.

1. はじめに

一般的な自然言語処理は、形態素解析、構文解析、意味解析といった解析を基礎に行われる。このうち意味解析は、解析結果を形態素解析結果や構文解析結果に反映させて各種曖昧性を解消したり、比喻理解や推敲の分野に利用したりするなどの有用性がある。

自然言語処理における意味解析には、語義の特定、深層格解析、その他が含まれるが、本論文ではこのうち深層格解析のみを対象として取り扱う。深層格解析とは述語(主に動詞)を文の中心に据え、その周りにある述語以外の構文要素に述語との意味的關係(深層格)を割り付けるといったものである^{1),2)}。

従来の深層格解析では、格フレームを用いる方法が最も一般的であった。格フレームとは、ある述語とそれがとりうる格とを表したパターンのことである。た

例えば、動詞 break の格フレームは次のようになる³⁾。

break - 対象格 [道具格] [動作主格]

ここで、対象格は動詞 break と義務的に共起する深層格(必須格)を意味し、[道具格]および[動作主格]は任意的に共起する深層格(任意格)を意味する。

格フレームを用いる深層格解析のうち代表的なものとしては、「意味素性付きの格フレームを用いる方法」と「用例付きの格フレームとシソーラスとを用いる方法」の2つがある。

「意味素性付きの格フレームを用いる方法」では、まず各名詞に意味素性(原子的な意味を表すシンボル)を割り付けておく。また各述語の格フレームには、あらかじめその述語と共起しうる格要素の深層格、表層格、およびその格要素が備えているべき意味素性、の3つ組を列挙しておく。この格フレームからの情報と、入力文の構文要素の表層格および意味素性とを比較することで、各構文要素に割り当てるべき深層格を決定する。

一方「用例付きの格フレームとシソーラスとを用いる方法」では、述語格フレーム中に、その述語と共起

[†] 九州産業大学工学部

Faculty of Engineering, Kyushu Sangyo University

^{††} 電子技術総合研究所

Electrotechnical Laboratory

しうる格要素の深層格、表層格、そして具体的な用例（単語）、の3つ組を記述しておく。またこれとは別に、単語間の意味的關係を階層的に表した上位下位シソーラスを用意する。システムは入力文内の構文要素と述語格フレーム内の各用例との意味關係をシソーラスによって調べ、その結果および各構文要素の表層格に基づいて深層格の認定を行う。

しかしながら、格フレームを用いたこれらの方法には次のような問題がある：

- (A1) 格フレーム作成のために人手を必要とするため、多大な労力が要求されるうえ、首尾一貫性を保つことが難しい。
- (A2) 実世界におけるそれぞれの語は、少しずつ重なりあいながら連続的に変化していくような意味的關係を持っているが、意味素性を用いる場合は、この連続的性質を（比較的少数の）離散的な集合で近似せざるをえない。このことは、意味素性に限らず、意味を「シンボル」で記述するときには避けられない問題である。
- (A3) シソーラスを利用する場合、シソーラス固有の問題——すなわち観点が1次元的であり⁴⁾、構造が静的である⁵⁾——がともなう。

上記の3つの問題を解決するには、次のような能力が必要である。(B1)~(B3)のそれぞれが(A1)~(A3)に対応する。

- (B1) コーパスから自動的に格フレームに相当する知識を獲得する学習能力。そして、その獲得した知識を未知の文に適用する一般化能力。
- (B2) 意味の距離を連続的に表現する能力。
- (B3) 多次元的で動的な観点に対応する能力。

このような能力は、ニューラルネットが得意とするところである。(B1)~(B3)に対応させて記述すると、ニューラルネットには(C1)~(C3)のような特長がある。

- (C1) 入力と出力のセット（ここでの場合、文と格構造のセット）に対する学習能力および一般化能力を有する。
- (C2) 入出力値として実数を取り扱うので、連続的な表現を扱うことができる。
- (C3) 学習データとして多次元的な観点を有するデータを与えることで、多次元的な観点を持つ処理を実現することができる。また、(C1)の学習能力ゆえに観点の動的な変化にも追従可能である。

そこで本論文では、ニューラルネットを用いた深層格解析について述べる。2章では、今回我々が提案するネットワークの基となった理論と、それを利用した

ネットワーク CAN について述べる。続く3章では今回提案するネットワークである CANAL の構成と、これを用いた格解析技法について述べる。4章ではネットワークの動作を確かめるためのシミュレーション実験とその結果について述べる。最後に5章で、実験結果と CANAL の問題点について、いくつかの考察を述べる。

2. 背景と原理

本章では、今回我々が提案するネットワークの基となった理論と、それを利用したモデルについて述べる。

Miikkulainen らは通常の誤差逆伝播法を拡張し、各ユニット間の結合荷重のみならず、入力表現をも、誤差が小さくなる方向に変化させていくアルゴリズム FGREP (Forming Global Representations with Extended backPropagation) を提案した^{6),7)}。概略は以下のとおりである。

一般的なニューラルネットにおける入力層ユニットの入出力関数は恒等関数である。つまり入力として受け取った値はそのまま出力される。恒等関数の微分は1であるので、一般化デルタルール⁸⁾より、その誤差信号は

$$\delta_i = \sum_j \delta_j w_{ij} \quad (1)$$

のようになる。ここで、 δ_i は入力層のユニット i に対する誤差信号、 δ_j は第1層目の中間層のユニット j に対する誤差信号、そして w_{ij} はユニット i とユニット j の間の結合荷重を表す。

通常の誤差逆伝播アルゴリズムではこの結合荷重を、次式の Δw_{ij} で表される差分だけ更新していく。

$$\Delta w_{ij} = \eta \delta_j r_i \quad (2)$$

ただし η は学習係数、 r_i は入力パターンの i 番目の値である。

FGREP ではさらに入力パターンを次式で示す Δr_i だけ修正する。

$$\Delta r_i = \zeta \delta_i \quad (3)$$

ここで ζ は更新係数である。

Miikkulainen らはこの FGREP に基づいたニューラルネットワークに英語の深層格解析を学習させた。入力となる英単語はベクトルの形で表現された（以下ではこのベクトルのことを単語表現ベクトルと呼ぶ）。単語表現ベクトルの初期値はランダムに設定されたが、深層格解析の学習が進行するに従って、類似した使われ方をする単語の表現ベクトルどうしが、次第に類似したベクトルへと収束することが示された。

Takahashi らは Miikkulainen らのニューラルネット

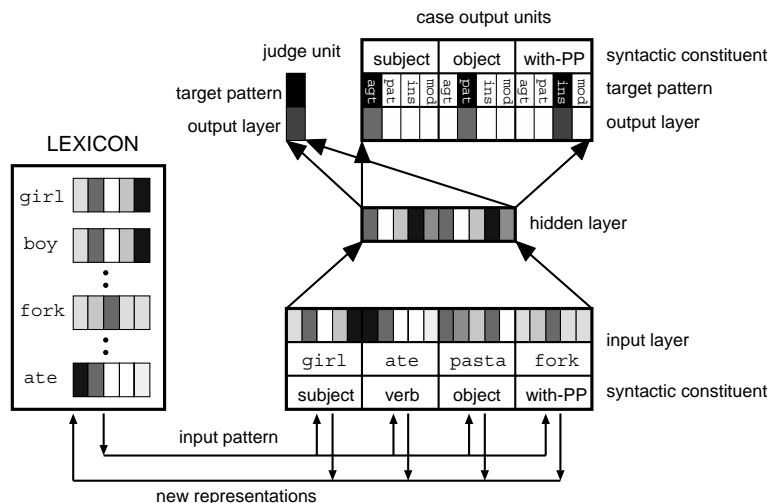


図1 CANALのネットワーク構成
Fig. 1 Network structure of CANAL.

トワークにおける出力表現を見直して効率化し、それをCAN (Case-role Assignment Network) と名付けた。CANはMiikkulainenらのネットワークに比べて、1) 全単語表現ベクトルが同一にならないことが保証される、2) 学習速度が速い、3) 学習が直接的である、という点が特長としてあげられる。彼らはまた、CANが一般化能力を備えていること、すなわち未学習の入力文に対しても深層格解析が可能であることを、クロスバリデーションテストを通じて示した^{9),10)}。

しかしながら、CANは深層格解析の学習を通じて単語表現ベクトルを獲得することを第1の目的としており、意味的に不適格な文が入力される可能性についてはまったく考慮していない。したがって、意味的に不適格な文をそうでない文から判別することが必要とされるような場面では用いることができない。

意味的に不適格な文をそうでない文から判別する機能は、推敲や比喩理解の分野で重要な役割を果たすと考えられる。そこで我々は、CANの基本的な機能に、文の選択制限性を判定する(すなわち意味的に不適格な文を検出する)機能を付加したニューラルネットワークモデルCANALを考案した。次章ではCANALの構成と、それをういた深層格解析の機構について説明する。

3. CANAL

3.1 ネットワーク構成

前章で述べたとおり、今回我々が提案するネットワークCANAL (Case-role Assignment Network for semantic AnaLysis) は、Takahashiらの提案したニュー

ラルネットワークモデルCANを基本としており、それとほぼ同様の構造を持つ。CANALは通常の(すなわちフィードバックループを持たない)階層型ニューラルネットワークの一種であり、その構成は図1に示すとおりである。

入力層は4つのスロットに分かれ、それぞれは入力文の4種類の構文要素(表層格)である主語、動詞、目的語、with前置詞句に対応付けられる。各スロットには、入力文を構成する各単語に対応する単語表現ベクトルが入力される。すなわち、ある入力文が与えられると単語表現ベクトルが4つつながった1つのベクトル(以下では、このベクトルのことを文ベクトルと呼ぶ)が形成され、これがCANALへと入力されることになる。なお入力文が1個以上の構文要素を欠く場合、そこに対応するスロットには“NULL”を表す特別なベクトルを入力する。

中間層は1層で、中間層ユニットの入出力関数 f_h は次式で示されるロジスティック関数である。

$$f_h(x) = \frac{1}{1 + e^{-x}}. \quad (4)$$

なお、中間層ユニット数が学習中に変化することはない。

出力層には入力文の主語、目的語、with前置詞句に対応する3つのスロットがあり、各々に対応する入力語の深層格を出力する。図1の例では、主語スロットは動作主格(agt)を、目的語スロットは対象格(pat)

動詞に対する深層格は考慮する必要がないので、動詞の出力スロットは設定しない。

を, with 前置詞句スロットは道具格 (ins) を出力している. 以下, これらの格役割を示す出力層ユニットをまとめて格出力ユニットと呼ぶ.

さらに入力文に対し深層格の格役割が認定可能であるかどうか, すなわち入力文が選択制限性を満たすかどうかを判定する 1 ユニットを付加する. これを判定ユニットと呼び, CANAL の独自性はここにある.

中間層ユニットと異なり, 出力層ユニットは入力値の総和をそのまま出力する. すなわち出力層ユニットの入出力関数 f_o は恒等関数である.

$$f_o(x) = x. \quad (5)$$

これは, 出力値が $[0.0, +1.0]$ の範囲内に収まる必要がないことと, 恒等関数はロジスティック関数に比べ計算量が少ないという 2 点を考慮した結果である.

入力層と中間層, および中間層と出力層の間はそれぞれ全結合されている. また入力層と出力層の間, および同一層内には結合は存在しない.

ネットワーク構成上で CANAL が CAN と異っているのは, 出力層に判定ユニットが追加されている点だけである. これは一見些細なことに見えるが, この追加によって CANAL は意味的に適切な文だけでなく, 意味的に不適格な文も取り扱うことが可能になっている.

3.2 学習手順

学習アルゴリズムは CAN と同様, FGREP を採用する. ただし, 選択制限性を満たさない文の学習に際しては, 判定ユニットに対してのみ教師信号との誤差を計算し, 格出力ユニットにおける誤差は無視する. したがって, 選択制限性を満たさない場合の誤差信号は次のようになる.

$$\delta_j = \delta_k w_{jk}. \quad (6)$$

ただし, δ_j は中間層のユニット j に対する誤差信号, δ_k は判定ユニット k に対する誤差信号, そして, w_{jk} はそれらの間の結合荷重を意味する.

なお, 選択制限性を満たす文に対しては, 通常どおりすべての出力層ユニットにおける誤差信号を考慮する.

3.3 深層格解析の機構

学習終了後の CANAL を用いた格解析の手順を示す. すべての入力文は構文的に正しいが, 意味的には正しいものとそうでないものとの両方が含まれる. まず, 与えられた入力文を構成する各単語の単語表現ベクトルを入力層から入力し, 前方向伝播の計算をする. 出力層の判定ユニットが発火した場合, すなわち入力文が意味的に正しいと判定された場合は, 格出力ユニット群の発火状態が各構成要素に割り当てられた

深層格のパターンを表す. 一方, 判定ユニットが発火しなかった場合, その入力文は選択制限性を満たさない (つまり意味的に正しくない) と判断されたことになる.

再び図 1 を例にとる. ここでは

A girl ate pasta with a fork.

という文を与えているが, 判定ユニットが発火していることから, 深層格の認定が可能 (すなわち意味的に妥当) と判断されていることになる. 同時に格出力ユニット群の発火状態から, 主語 (girl), 目的語 (pasta), with 前置詞句 (fork) のそれぞれに対して動作主格, 対象格, 道具格という深層格が認定されていることが分かる.

4. 実験

本章では, CANAL による格解析実験とその結果について述べる.

4.1 コーパス

実験に用いたコーパスは, 計算機により生成した 57701 文である. 用いた単語は Takahashi ら^{9),10)} と同じ名詞 24 語, 動詞 4 語であるが, Takahashi らが選択制限性を満たす文のみを扱っているのに対し, 本論文ではそれに加えて選択制限性を満たさない文も処理対象とする. 以下では, 選択制限性を満たす文を正文, 選択制限性を満たさない文を逸脱文と呼ぶことにする.

正文の作成には, 表 1 に示すメタワードと表 2 に示す文テンプレートの組合せを用いた. すべての組合

表 1 メタワードと実際の単語の対応関係

Table 1 Correspondence between meta-words and actual words.

meta-word	nouns
HUMAN	man woman boy girl
ANIMAL	bat chicken dog sheep wolf lion
PREDATOR	wolf lion
PREY	chicken sheep
FOOD	chicken cheese pasta carrot
UTENSIL	fork spoon
FRAGILE	plate window vase
HITTER	bat ball hatchet hammer vase paperwt rock
BREAKER	bat ball hatchet hammer paperwt rock
POSSESSION	bat ball hatchet hammer vase dog doll
OBJECT	bat ball hatchet hammer paperwt rock vase plate window fork spoon pasta cheese chicken carrot doll
THING	HUMAN ANIMAL OBJECT
THING2	ANIMAL OBJECT

表 2 正文テンプレート
Table 2 Templates for acceptable sentences.

No.	input pattern				target pattern		
	subject	verb	object	with-PP	subject	object	with-PP
1	HUMAN	ate			agent		
2	HUMAN	ate	FOOD		agent	patient	
3	HUMAN	ate	FOOD	FOOD	agent	patient	modifier
4	HUMAN	ate	FOOD	UTENSIL	agent	patient	instrument
5	ANIMAL	ate			agent		
6	PREDATOR	ate	PREY		agent	patient	
7	HUMAN	broke	FRAGILE		agent	patient	
8	HUMAN	broke	FRAGILE	BREAKER	agent	patient	instrument
9	BREAKER	broke	FRAGILE		instrument	patient	
10	ANIMAL	broke	FRAGILE		agent	patient	
11	FRAGILE	broke			patient		
12	HUMAN	hit	THING		agent	patient	
13	HUMAN	hit	HUMAN	POSSESSION	agent	patient	modifier
14	HUMAN	hit	THING2	HITTER	agent	patient	instrument
15	HITTER	hit	THING		instrument	patient	
16	HUMAN	moved			agt+pat		
17	HUMAN	moved	OBJECT		agent	patient	
18	ANIMAL	moved			agt+pat		
19	OBJECT	moved			patient		

せから重複を削除した後の文数は 1275 である。

一方逸脱文は、表 1 に示すメタワードと表 3 に示す文テンプレートによって生成される全候補文から、上記の正文および重複したパターンを除くことで作成した。これら 56426 の逸脱文の中には、一般常識に照らして意味的に妥当と考えられる文も何文が含まれていた。しかし本実験では上述の正文のみが意味的に妥当であると“定義”し、その定義の下での CANAL の振舞いを調べることにした。

4.2 ネットワーク規模実験

まず、CANAL の学習能力および一般化能力が、中間層のユニット数によってどのように変化するか調べた実験について述べる。

この実験では中間層ユニットを 5 ユニットから 20 ユニットまで 5 ユニットずつ変化させ、それぞれのネットワークで RMS (root mean square) 誤差が 0.04 になるまで最大 100 万回の学習を行った。学習が終了した時点で、学習コーパスおよびテストコーパスに対する格解析テストの正答率を調べた。

ただし今回使用したコーパスに含まれる異なり単語

なお、この 1275 文の中には入力パターンが同一で目標パターンが異なる例が 5 組存在する。すなわち、入力パターンだけに限れば異なり文数は 1270 となる。

たとえば、“A chicken ate pasta with cheese.” という文は一般常識に照らして意味的に妥当と考えられるが、ここでの定義に従えば逸脱文となる。

表 3 全候補文テンプレート。THING は前述のメタワード

Table 3 Templates for all possible sentences. THING is the meta-word previously appeared.

No.	input pattern			
	subject	verb	object	with-PP
1	THING	ate		
2	THING	ate	THING	
3	THING	ate	THING	THING
4	THING	broke		
5	THING	broke	THING	
6	THING	broke	THING	THING
7	THING	hit		
8	THING	hit	THING	
9	THING	hit	THING	THING
10	THING	moved		
11	THING	moved	THING	
12	THING	moved	THING	THING

数および文数はどちらも十分といえないため、そのままでは実験結果に偏りが生じるおそれがある。これを避けるため、以下の要領で試行回数 10 回のクロスバリデーションテストを行った。

まず、正文(選択制限性を満たす文) 1275 文を 10 等分し、そのうちの 3 セット(平均 382.5 文)を学習コーパス、残り 7 セット(平均 892.5 文)をテスト用コーパスとした。一方、逸脱文(選択制限性を満たさない文) 56426 文は 1000 等分し、そのうちの 10 セット(平均 563.9 文)を学習コーパス、残りの 990 セット(平均 55862.1 文)をテストコーパスとした。また、学習コーパスの提示順序が学習に偏りを及ぼすことを

表 4 中間層ユニット数と学習速度および格解析正答率との関係
Table 4 Relation of the number of hidden units, learning speed and precision of analysis.

中間層 ユニット数 (ユニット)	平均 学習回数 (sweeps)	正答率 (%)					
		对学习データ			对テストデータ		
		正文	逸脱文	全体	正文	逸脱文	全体
5	未収束	-	-	-	-	-	-
10	21387	99.09	99.93	99.59	96.04	95.99	95.99
15	10612	99.42	99.96	99.75	96.38	96.31	96.31
20	11091	99.37	99.89	99.68	95.87	96.76	96.75

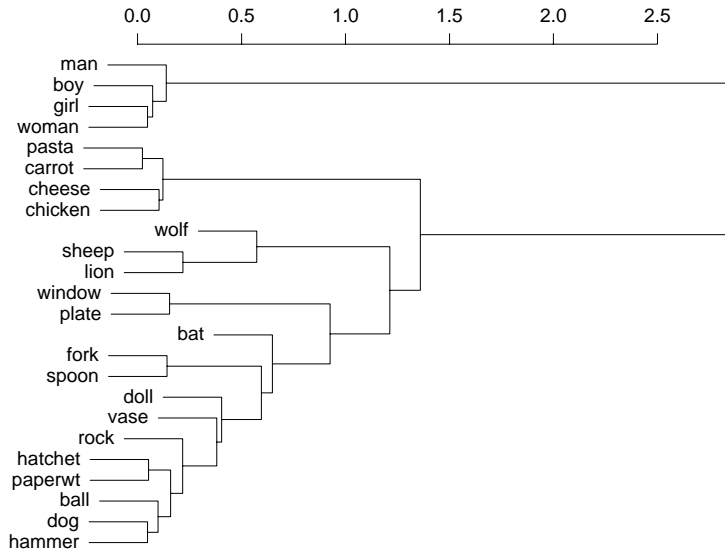


図 2 単語表現ベクトルのクラスター分析
Fig. 2 Cluster analysis result of the word representation vectors.

防ぐため、各学習に先立って学習コーパスを1度ランダムにシャッフルした。

学習におけるその他の条件は以下のとおりである：各単語表現ベクトルは5次元とした。「該当構文要素なし」を示す“NULL”ベクトルはすべての要素が1.0で、学習中も変化しない。単語表現ベクトルの各要素、ユニット間の結合荷重、およびユニットのバイアスはそれぞれ[-1.0, +1.0]の乱数で初期化した。格出力ユニットが発火した状態の目標出力値は0.9、発火していない状態の目標出力値は0.1とした。判定ユニットに関しては、発火した状態の目標出力値は1.0、発火していない状態の目標出力値は0.0とした。

学習には逐次学習法を採用し、1文の前方向の伝播処理を行うごとに、出力信号と教師信号の誤差を逆方向に伝播し、結合荷重とバイアス、そして単語表現ベクトルの更新処理を行った。結合荷重の学習率(式(2)の η)は0.01、慣性項はなしとした。また単語表現ベクトルについても、各要素の更新係数(式(3)の ζ)は0.01、慣性項はなしとした。結合荷重の値には制限を設けないが、単語表現ベクトルの各要素値は

[-1.0, +1.0]に収まるようにした。具体的には、時刻 t における入力パターンの i 番目の要素を $r_i(t)$ 、 Δr_i を式(3)の定義どおりとしたとき、

$$r_i(t+1) = \max(-1, \min(+1, r_i(t) + \Delta r_i)) \tag{7}$$

とした。

以上の処理を学習コーパス中の全入出力ペアに対して1回ずつ行うことを1回の学習(1 sweep)とした。

格解析テストとしての正答は次のように定義した。正文に対しては、判定ユニットが0.75より大きい値を示し、かつすべての格出力ユニットの値が対応する教師信号の値の ± 0.15 以内に収まっている場合を正答とした。また逸脱文に対しては、判定ユニットの値が0.25より小さい値を示したもの(格出力ユニットの値に関しては不問)を正答とした。

実験結果を表4にまとめた。この表から中間層ユニットが5ユニットでは少なすぎるが、10ユニットあればそれ以上の中間層ユニットを持つネットワークとほぼ同等の格解析能力を有するネットワークを作成できることが分かる。

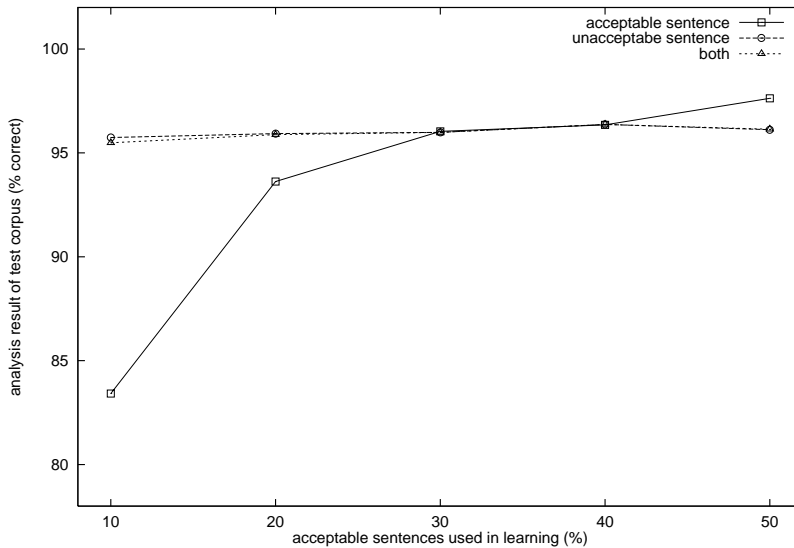


図 3 学習正文数と対テストコーパス正答率との関係

Fig. 3 Relation between number of acceptable sentences used in learning and analysis result of test corpus.

図 2 に、中間層ユニット数 10, RMS 誤差 0.04 の収束時における単語表現ベクトルのクラスタ分析図 (ユークリッド距離, 群平均法) を示す. この図から CANAL も他の先行研究と同様, 類似した使われ方をしている単語には類似した単語表現ベクトルを与えることが確認できる.

4.3 学習データ量と正答率に関する実験

次に, 学習データの量が正答率に与える影響を調べる実験を行った.

まず, 正文 1275 文を 10 等分し, そのうちの m セットを学習コーパス, 残り $(10 - m)$ セットをテストコーパスとした. 一方, 逸脱文 56426 文は 1000 等分し, そのうちの n セットを学習コーパス, 残りの $(1000 - n)$ セットをテストコーパスとした. そして, 以下の 2 つの実験を行った.

実験 (1): 学習データ中の正文の量が正答率に及ぼす影響を調べるため, 正文の m を 1, 2, 3, 4, 5 と変化させる実験を行った. 逸脱文の n は 10 (すなわち学習データとテストデータの量の比は 1 : 99) に固定した. すなわち, 作成した全正文のうち学習に用いるものの割合を 10% (127 文), 20%, 30%, 40%, 50% と変化させた.

実験 (2): 学習データ中の逸脱文の量が正答率に及ぼす影響を調べるため, 逸脱文の n を 2, 4, 6, 8, 10 と変化させる実験を行った. 正文の m は 1 (すなわち学習データとテストデータの量の比は 1 : 9) に固定した. すなわち, 作成した全逸脱文のうち学習に用いるものの割合を 0.2% (113 文), 0.4%, 0.6%, 0.8%, 1.0% と変化させた.

いずれの場合も, 10 回のクロスバリデーションテストを行って, 実験結果の偏りを防ぐようにした. ネットワークの中間層ユニット数は 10 とし, その他はすべて前節と同じ条件とした.

学習終了後, 格解析テストを行ったところ, 学習コーパスに対しては, すべての場合において正答率が 99.0% を上回る結果となった. また, テストコーパスに対する実験 (1) の結果は図 3 のとおり, 実験 (2) の結果は図 4 のとおりであった. 図 3 より, 学習させる正文の数を増加させるに従って, 正文に対する正答率が向上することが分かる. さらに図 4 より, 学習させる逸脱文の数を増加させるに従って, 逸脱文に対する正答率が向上することが分かる. またこの場合, 正文に対する正答率が低下する傾向にある. これは, 全学習データ中で逸脱文が占める割合が増大するに従って, 判定ユニットが発火しにくい方向に学習が進み, その結果学習されていない文 (すなわちテスト文) が逸脱文と判定される傾向が強くなるためと考えられる.

4.4 FGREP 効果確認実験

次に, FGREP による入力ベクトル更新が学習と解

動詞は第 2 入力スロットにしか与えられず, 反対に名詞は第 2 入力スロットには決して与えられない. つまり動詞の単語ベクトルと名詞の単語ベクトルとは, 2 つの異なるベクトル空間内に存在することになる. したがって, ここでは名詞ベクトルのみを示した.

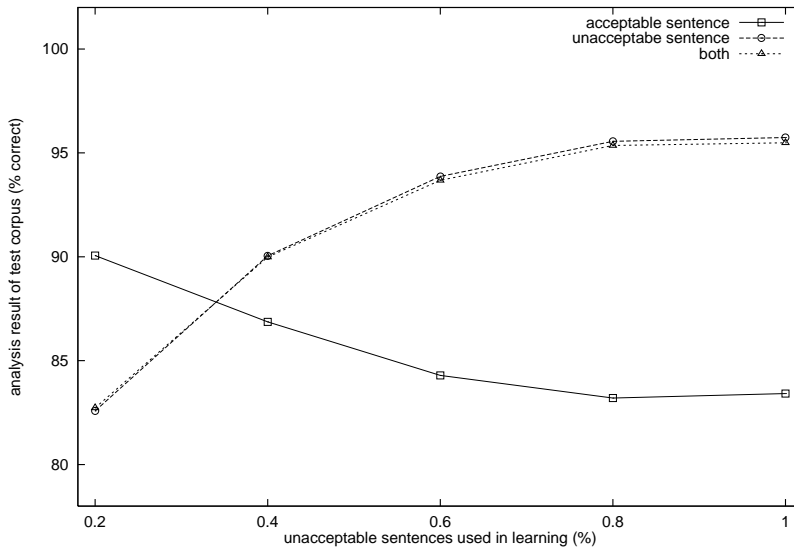


図 4 学習逸脱文数と対テストコーパスの正答率との関係

Fig. 4 Relation between number of unacceptable sentences used in learning and analysis result of test corpus.

析に有用であることを確認するための実験を行った。具体的には、

- FGREP に従い、学習時にユニット間結合荷重と入力ベクトルの両方を更新するネットワーク
- 通常の誤差逆伝播法に従い、ユニット間結合荷重のみを更新するネットワーク

の両者において中間ユニット数を同一に設定し、それぞれの学習速度および学習終了後の格解析正答率を比較した。どちらの場合も 1 ユニットあたりの RMS 誤差が 0.025 以下になるまで学習を続け、その後格解析テストを行った。

予備実験を行った結果、通常の誤差逆伝播法で中間層ユニット数が 70 以下の場合、学習を 20 万回まで繰り返しても RMS 誤差が 0.025 まで収束しない可能性のあることが分かった。収束を確実なものとするため、中間層ユニット数は通常ネットワーク、FGREP ネットワークとも 80 に設定した。両ネットワークにおける各種パラメータは、式 (3) の ζ 以外はすべて前節の実験と同一に設定した。

実験の結果は以下のとおりである。10 回のクロスバリデーションテストのうち最初の試行において、学習中の RMS 誤差がどう変化するかを図 5 に示す。この図から FGREP 法を用いた場合の方が、通常の誤差逆伝播法を用いた場合より学習の進行が速いことが

分かる。

また、両者の学習が収束した後に実施した格解析テスト結果を表 5 に示す。この表に示されるように、FGREP 法を用いた場合、通常の誤差逆伝播法を用いた場合と比較して、学習データに対するテストでは 0.07%、テストデータに対するテストでは 31.01%、全文に対するテストでは 30.50%、それぞれ正答率が上回った。

5. 考 察

本章では、実験結果と CANAL の問題点について考察する。

5.1 実験結果の検討

4 章の実験結果から、人工コーパスを学習対象とした場合において CANAL は深層格解析の学習が可能であり、また満足できる一般化能力を有することが分かった。また、学習回数 20 万回以内で RMS 誤差が 0.025 になるまで学習させようとした場合、通常の誤差逆伝播法では 80 個程度の中間層ユニットが必要だったが、FGREP 法を用いた場合はこれが 10 個程度で十分であった。さらに一般化テストの正答率は、FGREP 法を用いた場合が通常の誤差逆伝播法を用いた場合を 31.01% 上回っていた。これらのことから、FGREP による単語表現ベクトルが、1) ネットワークサイズの削減(すなわち計算コストの削減)と、2) 格解析能力の

これは式 (3) の ζ を 0 に設定することと等価である。このとき、各単語表現ベクトルは、学習開始時に与えられたランダムベクトルのまま固定されることになる。

他の 9 回の試行もほぼ同様の結果となった。

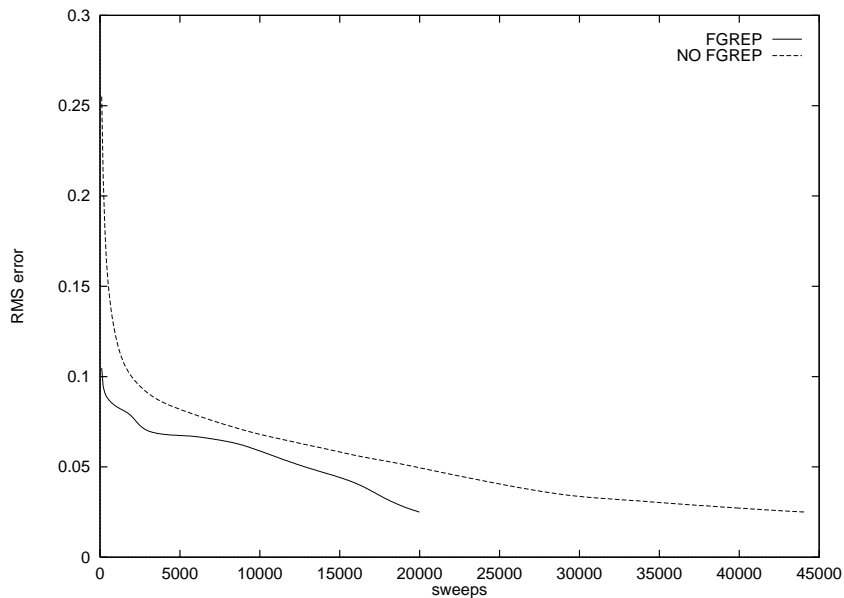


図 5 学習過程における RMS 誤差の変化

Fig. 5 Behavior of RMS error during a learning session.

表 5 学習方法と格解析テスト結果の関係

Table 5 Relation between training method and precision of analysis.

学習方法	正答率 (%)						全文
	对学习データ			对テストデータ			
	正文	逸脱文	全体	正文	逸脱文	全体	
通常の誤差逆伝播法	99.69	100.0	99.87	60.01	64.88	64.80	65.38
FGREP	99.84	100.0	99.94	95.63	95.81	95.81	95.88

向上, の 2 点に貢献していることが確かめられた.

格解析能力向上の理由は, 次のように説明できる. 今, 文 (a) 含み, かつ文 (b) を含まないようなコーパスを CANAL に学習させたとする.

(a) A girl ate pasta with a fork.

(b) A boy ate pasta with a fork.

コーパスには (a) と (b) 以外にも多くの文が含まれているはずであり, そこでは単語 girl と単語 boy とは比較的——たとえば girl と bicycle の関係などに比較すると——類似した使われ方がなされていると考えられる. ゆえに girl の単語表現ベクトルと boy の単語表現ベクトルは類似したベクトルとなっているはずである. また文 (a) と文 (b) の違いは主語が girl であるか boy であるかだけなので, もし girl の単語表現ベクトルと boy の単語表現ベクトルが類似しているならば, 文 (a) の文ベクトルと文 (b) の文ベクトルも類似していることになる.

さて, ニューラルネットの実行処理 (前方向の伝播) は, 多次元空間内の入力分布に対する非線形分離機能を基本としているので, 類似した 2 つの入力に対し

ては類似した値が出力される. 文 (a) は学習コーパス中に含まれているのであるから, CANAL は文 (a) の文ベクトルが入力された場合は正しい格解析結果を出力するように学習を終了させているはずである. したがって, 未学習の文 (b) の文ベクトルが入力された場合も, それが文 (a) の文ベクトルと類似しているために正しい格解析結果が出力されることになる.

以上は, 入力文が正文である場合の一般化能力についての考察だが, 入力文が逸脱文である場合についても同様の議論が成り立つ. したがって, コーパスが望ましい例文を備えているならば, CANAL は未学習の文に対しても一般化能力を発揮すると予想される.

なお, 図 2 は Takahashi ら⁹⁾ の CAN に対する分析結果に比べるといくぶんまとまりが悪いように見える. たとえば, 単語 'dog' は, CAN では 'sheep' や 'lion' の近くに位置しており, 動物のカテゴリにあったが, CANAL では, 'hammer' など道具のカテゴリの近くに位置している. これは, CAN の学習が正文のみであったのに対し, CANAL の学習は逸脱文を含んでいたためであると考えられる.

逸脱文の学習のために、獲得した単語表現ベクトルがまとまりの悪いものになったのであれば、CANALの独自性である逸脱文の学習が悪い影響を与えることになる。この悪影響をおさえるためには、FGREPによる単語表現ベクトルの学習を正文のみで行うことが考えられる。

そこで、入力が正文の場合のみ単語表現ベクトルの更新を行うように学習方法を変えて追加実験を行った。実験条件として、学習コーパスとテストコーパスの比は、正文に関しては3:7、逸脱文に関しては1:99とした。また、中間層ユニット数を10, 40, 60, 80, 100, 200, 300と変化させてみた。しかしこれらのネットワークでは学習がうまく進まず、RMS誤差が0.04まで減少することはなかった。

以上の追加実験の結果から、学習方法を、入力が正文の場合のみ単語表現ベクトルの更新を行うようにすることは難しい、ということがいえる。

5.2 CANALの問題点

FGREPの利用により、CANALは通常の階層型ニューラルネットよりも少ない中間層ユニット数で同程度の性能の格解析を実現できることが分かった。

しかしながら、我々が認識しているだけでもCANALには以下のような問題点がある。

(1) 拡張性(スケーラビリティ)

今回の実験対象は、単語数が28という非常に小規模なものである。FGREPによりかなりの効率化が実現できているものの、果たしてどれだけの規模の単語、文までカバーしうるのかは解明されていない。

(2) データの過疎性(データスパースネス)

今回の実験のような、生成ルールを用いて作成した小規模なコーパスに関しては、選択制限知識の学習と一般化はほぼ成功したといえる。しかし、4.3節の実験では、学習コーパスの量が小さい場合は、十分な量が確保されている場合に比べ、一般化能力が劣ることが明らかとなった。実コーパスにおいては、学習コーパスとテストコーパスの比率は、今回用いた人工コーパスにおける比率とはかなり異なると思われるが、データの過疎性が、学習コーパスと適用可能なテストコーパスの比率に影響を及ぼす程度は、まだ分かっていない。

(3) 学習方法論の構築

格解析実行時の計算量は少ないが、それに先立つ学習時の計算量は学習コーパスの量とともに増大する。また学習済みのネットワークに新たな用例を追加学習させる場合、単に新たな用例を学習させるだけでなく、元の学習コーパスの再学習も必要である。なぜなら新たな用例だけを追加学習させた場合、それによる変化

が以前の学習内容に悪影響を与える可能性があるからである。さらに、どのような文を学習コーパスに選べば一般化の効率が良いのか、といった研究も必要となるであろう。

(4) 未知語への対応

CANALは、未学習の単語を含む文を処理することができない。1つの対策案として、Takahashiらは、学習済みの単語表現ベクトルと上位下位シソーラスの情報から未学習の単語表現ベクトルを推定する方法を提案している⁹⁾が、この手法の有効性は実験によって実際に確認されたわけではない。

(5) 辞書と解析器のモジュール化

CANALによって獲得された単語表現ベクトルは、CANAL内の結合荷重と密接に結び付いている。そのため、単語表現ベクトルとニューラルネットの結合荷重を切り離して個別に利用することができない。このことはシステム化に際して、単語表現ベクトルの拡張と、格解析性能の向上とがそれぞれ独立して行えないことを意味する。

(6) 逸脱文の収集

学習コーパスとして選択制限性を満たす文を用意することでさえデータスパースネスの問題が存在しているのに、選択制限性を満たさない文を用意するとそのような文のデータスパースネスの問題も存在することになる。考えられる1つの策として、現在稼働中のシステムにおいて意味解析の結果不適格とされた文を人間の手により判別し、選択制限性を満たさない文を収集するような半自動ツールの作成があげられる。

以上、CANALの問題点を列挙した。これだけの問題点を考えると、CANALの実用システムへの適用はまだ難しいように思える。しかし、(1)と(2)、および、(3)と(4)と(5)、はそれぞれ互いに密接な関係を持つ問題である。したがって、1つのアイデアや研究結果が上記の問題の2つ以上に対して解決を与える可能性も考えられる。これらについては、稿を改めて述べたい。

6. おわりに

本論文で提案した階層型ニューラルネットによる深層格解析は、いうなれば例文より自動的に解析のアルゴリズムと辞書データを獲得する方法である。そこで要求される作業のうちで人手が必要なものは、ネットワークに学習させるべき例文の収集と選択、および各例文の構文要素に対する格役割の付与のみとなる。これまでの人手を用いた格フレームの構築は、その編集者の技量や主観的決断にも負うところが大きいと考え

られる．本研究が進むことによって，このような手作業はより軽減されるであろうと予想する．

謝辞 自然言語処理に関する助言をいただいた，九州産業大学工学部電気工学科の藤田毅助教授，久留米工業大学の髙橋雅仁助教授に感謝いたします．また，ニューラルネットに関して助言をいただいた九州工業大学情報工学部石川眞澄教授に感謝の意を表します．

参 考 文 献

- 1) Fillmore, C., 田中春美, 船城道雄(訳): 格文法の原理, 三省堂 (1975).
- 2) 水谷静夫(編): 朝倉日本語新講座 3 文法と意味 I, pp.81-134, 朝倉書店 (1983).
- 3) 長尾 真(編): 岩波ソフトウェア科学講座 15 自然言語処理, pp.199-230, 岩波書店 (1996).
- 4) 川村和美, 片桐康祐, 宮崎正弘: 語を種々の観点から分類した多次元シソーラス, 電子情報通信学会 言語理解とコミュニケーション研究会報告, pp.33-40 (1995).
- 5) 浦本直彦: コーパスに基づくシソーラス-統計情報を用いた既存のシソーラスへの未知語の配置, 情報処理学会論文誌, Vol.37, No.12, pp.2182-2189 (1996).
- 6) Miikkulainen, R. and Dyer, M.: Natural Language Processing with Modular PDP Networks and Distributed Lexicon, *Cognitive Science*, Vol.15, pp.343-399 (1991).
- 7) Miikkulainen, R.: Subsymbolic Parsing of Embedded Structures, *Computational Architectures Integrating Neural and Symbolic Processes*, Sun, R. and Bookman, L.A. (Eds.), pp.153-186, kluwer academic publisher (1996).
- 8) Rumelhart, D., Hinton, G. and Williams, R.: *Learning Internal Representations by Error Propagation*, Vol.1, The MIT Press, Massachusetts (1986).
- 9) Takahashi, N. and Motoki, M.: A subsymbolic approach for acquiring semantic representations, *3rd International Workshop on Computational Semantics* (1999).
- 10) 髙橋直人: ニューラルネットを用いた意味表現形式の自動獲得, 電子情報通信学会言語理解とコミュニケーション研究会報告, pp.17-24 (1998). (平成 11 年 12 月 3 日受付)
(平成 12 年 9 月 7 日採録)



本木 実(正会員)

1970 年生．1995 年熊本大学大学院工学研究科電気情報工学専攻生体情報処理講座修士課程修了．同年沖電気工業(株)入社．陸上自衛隊指揮システムの開発に従事．1997 年より九州産業大学工学部電気工学科研究生．現在九州産業大学工学部電気工学科助手．認知科学，および自然言語処理研究でのシンボリズムとコネクショニズムの協調的システムに興味を持つ．電子情報通信学会，言語処理学会各会員．



嶋津 好生(正会員)

1963 年九州大学工学部電子工学科卒業．1967 年同大学大学院工学研究科電子工学専攻博士課程中途退学．同年九州産業大学工学部電気工学科常勤講師赴任．1987 年同教授昇任，現在に至る．工学博士(1986 年九州大学)．主としてコネクショニズムに基づく自然言語処理，認知科学の研究に従事．日本認知科学会，言語処理学会，人工知能学会，日本神経回路網学会各会員．



髙橋 直人(正会員)

1987 年筑波大学第三学群情報学類卒業．1992 年同大学大学院博士課程工学研究科修了．博士(工学)．同年電子技術総合研究所入所．1996 ~ 1997 年フランス LEIBNIZ-IMAG 研究所客員研究員．自然言語処理，ニューラルネットワーク，マルチリンガル情報システム等の研究に従事．人工知能学会，言語処理学会各会員．