

ニューラルネットを用いた句のベクトル表現*

電子技術総合研究所

高橋直人

ntakahas@etl.go.jp

九州産業大学

本木実

motoki@te.kyusan-u.ac.jp

1 はじめに

Miikkulainen は FGREP と呼ぶニューラルネットを用い、コーパスから単語のベクトル表現を自動獲得することに成功した [1]。

彼はまずバックプロパゲーションを拡張し、ニューラルネット内のリンクの重みだけでなく、入力信号をも更新する方法を考案した。基本的な原理は以下の通りである。

ニューラルネットにおける入力ユニットは他のユニットとは異なり、外部から与えられた値をそのまま次の層へと送るだけである。すなわち入力ユニットは、恒等関数を入出力関数とする計算ユニットと考えることができる。

一般に、出力ユニット以外のユニットにおける誤差関数の式は、

$$\delta_i = f'_i(\text{net}_i) \sum_j \delta_j w_{ij} \quad (1)$$

で与えられる。ただし i を j より入力に近い側の層に属するユニットとしたとき、 net_i はユニット i における全入力、 f'_i はユニット i の入出力関数、 δ_i はユニット i における誤差信号、 w_{ij} はユニット i とユニット j を結ぶリンクの重みである。[2]。

したがって入力ユニットの誤差関数は、上式の f' に恒等関数 $f(x) = x$ の微分である 1 を代入した

$$\delta_i = \sum_j \delta_j w_{ij} \quad (2)$$

となる。よって入力ベクトルの変更は

$$\Delta r_{ci} = \eta \delta_i \quad (3)$$

に従って行えばよい。ここで r_{ci} はベクトル c の第 i 要素、 η は学習の係数である。

*Representing phrases as vectors using neural networks by TAKAHASHI Naoto and MOTOKI Minoru:

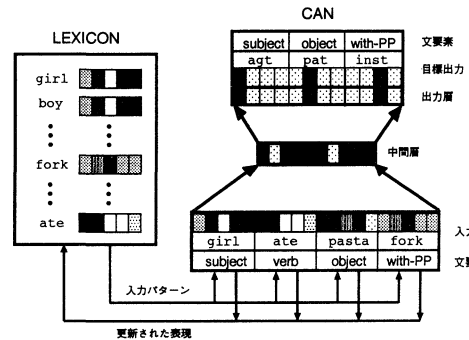


図1: CANの構成図。学習すべきタスクは、入力文の各構成要素に対してその格役割を割り当てることである。なお、動詞の格役割は常に act であるので出力は省略される。

FGREP による単語のベクトル表現獲得は、まず各単語に初期値としてランダムなベクトルを割り当て、次に上記の拡張されたバックプロパゲーションによって単語ベクトルを徐々に書き換えることによって行われた。学習すべきタスクは、入力文の各構成要素（主語、動詞、目的語、with-前置詞句）に格役割を付与するというものである。入力文としては人工的に作製した小規模のコーパスに含まれるものが用いられた。学習が進むにつれ、コーパス中で互いに似た用法を持つ単語のベクトル同士は次第に近付き合い、全体としていくつかのクラスタを形成するようになった。

著者らは FGREP の目標出力パターンに変更を加えて学習を高速化すると共に、FGREP が潜在的に持ついくつかの問題点を解決することに成功した [3][4]。我々が用いたニューラルネット CAN の構成と、それによって獲得された単語ベクトル表現のクラスタリング結果を、それぞれ図1と図2に示す。

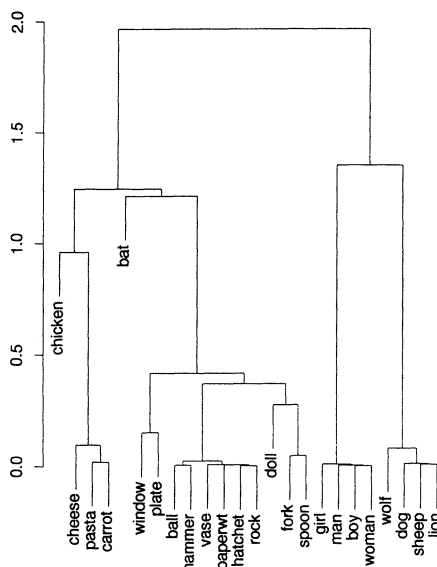


図 2: CAN によって獲得されたベクトル表現を、ユークリッド距離によって分類した結果。コーパス中で似たような使われ方をする単語ほど似たようなベクトルを持っていることがわかる。Chicken には動物と食物の 2 つの用法が、また bat には動詞と野球のバットの 2 つの用法があるので、他の語とは離れて独立した存在となっている。

2 名詞句ベクトルの合成

CAN も FGREP と同様、入力文の各格要素は単純な語(名詞)に限られ、2 語以上からなる名詞句をとることはできなかった。我々は新たなニューラルネット 2 つを CAN と協調動作させることで、入力文の格要素として「形容詞+名詞」の形の名詞句が許されるようにした。¹

2.1 Composer ネットワーク

新たに導入された 2 つのニューラルネットのうち第 1 のものは composer と呼ばれ、

- 名詞と形容詞から名詞句ベクトルを合成
- 形容詞ベクトルの獲得

¹本稿では「形容詞+名詞」型の名詞句についてのみ述べるが、この手法は他の形の名詞句にも応用可能である。

という 2 つの役割を持つ。CAN と composer を用いた学習は、概略以下のように進む。(図 3)

まず格要素が名詞 1 語の場合、その名詞のベクトル表現は直接 CAN に入力される。CAN によって計算された出力と目標出力との誤差は拡張されたバックプロパゲーションによって CAN の中を逆伝搬し、リンクの重みおよび入力された名詞のベクトル表現を更新する。図 3 中では、girl および hammer がこれに当たる。

次に格要素が「形容詞+名詞」の場合について述べる。図 3 中では文の目的語である big window がこれに当たる。このときはまず形容詞と名詞それぞれのベクトル表現を composer に入力して新たなベクトルを合成し、次にその合成されたベクトルを格要素として CAN の入力層に渡す。CAN の出力層における誤差信号は CAN の入力層を経てさらに composer の中を逆伝搬する。この信号によって 1) CAN のリンクの重み、2) composer のリンクの重み、3) composer の入力層における形容詞のベクトル表現、の 3 つを書き換える。

なお composer の入力層において書き換えられるのは形容詞ベクトルだけであり、名詞ベクトルは変化を受けない。名詞ベクトルが書き換えられるのは CAN の入力層においてのみである。

2.2 Decomposer ネットワーク

新たに composer を追加したことによって、「形容詞+名詞」型の名詞句も処理できるようになったが、これだけでは以下のような好ましくない問題の生じる可能性がある。

後述するように、今回の実験では入力文の格要素に形容詞が付いても付かなくても主辞である名詞の格役割は変化しないと仮定している。すなわちある名詞と、その名詞に形容詞が付いた名詞句とは、コーパス中においてまったく同じように使われることになる。このことは、ある名詞のベクトル表現と、その名詞に形容詞が付いた名詞句のベクトル表現とが同一のものになる可能性が高いことを示唆している。もしある名詞のベクトル表現と、その名詞に形容詞が付いた名詞句のベクトル表現とが常に同一のものになるならば、composer の学習は入力層の形容詞部を無視し、単に名詞部のコピーを出力する方向に進むと考えられる。

これを防ぐためには、合成された名詞句ベクトルに、入力された名詞ベクトルと形容詞ベクトルの両方の情報が必要含まれるような制約を加える必要がある。そのため 1 つの方法は、合成された名詞句ベクトルから、元の名詞ベクトルと形容詞ベクトルの両方を復元するとい

入力文 :
The girl broke the big window with the hammer.

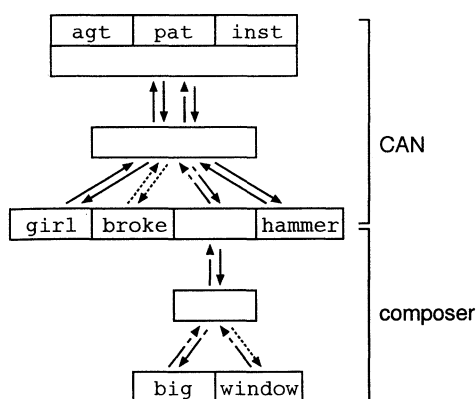


図 3: 格要素が名詞 1 語の場合と、名詞句である場合との違い。名詞 1 語の場合、そのベクトル表現は直接 CAN に入力され、CAN の入力部で変形を受ける (実線)。一方名詞句の場合は、まず composer によって形容詞ベクトルと名詞ベクトルが合成され、その結果が CAN に渡される。誤差信号は CAN の入力層を経て composer に戻り、composer の入力層では形容詞ベクトルのみが変形を受ける (破線)。

うタスクを与えることである。このタスク実現のため、もう 1 つのニューラルネット decomposer を導入した。

各入力文に関する composer と CAN の協調学習の後には、composer と decomposer との協調学習が行われる。このとき composer の出力は decomposer の入力に結び付けられ、両者は合わせて 5 層ニューラルネットとなる。ここでのタスクは auto-association である (図 4)。Decomposer の出力層における誤差信号は decomposer, composer の順に逆伝搬され、composer の入力層に与えられた形容詞のベクトルを書き換える。ここでもまた、名詞ベクトルは書き換えられない。

以上整理すると、文の格要素として名詞句が与えられた場合、composer 内部のリンクおよび形容詞ベクトルは、CAN に由来する誤差信号と、decomposer に由来する誤差信号の両方によって修正されることになる。

3 実験

[3, 4] で用いた学習用例文 (文パターン 19、異なり語数 29、異なり文数 1275) に big, small, old, young,

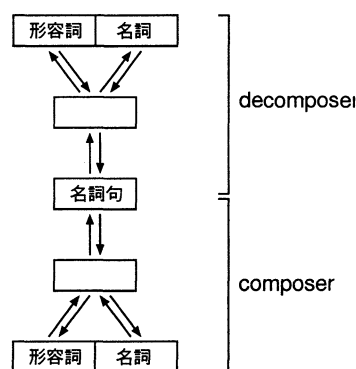


図 4: Composer と decomposer の協調学習。2 つのニューラルネットは直結されて 5 層ニューラルネットとなり、入力に与えられた 2 つの単語ベクトルを出力層に再現するというタスクが与えられる。

cheese, pasta, carrot :
常に形容詞なし
boy, girl :
tall, short それぞれ 0.2; 形容詞なし 0.6
man, woman :
tall, short, old, young, 形容詞なし それぞれ 0.2
dog, sheep, wold, lion :
big, small, old, young, 形容詞なし それぞれ 0.2
その他の名詞 :
big, small それぞれ 0.2; 形容詞なし 0.6

表 1: 入力文中の各単語が伴う可能性のある形容詞とその出現確率

tall, short という形容詞 6 語を追加し、CAN, composer, decomposer の協調学習を行った。各名詞が形容詞を伴って現れる確率は表 1 に示す通りである。また前述のように、入力文の格要素に形容詞が付いても付かなくても主辞である名詞の格役割は変化しないと仮定している。入力文はランダムに選んだ文パターンをもとに動的に生成し、のべ 10 万文の学習をもって 1 セッションとした。

ニューラルネットの中間ユニット数は CAN が 10、composer と decomposer は 5 である。いずれのニューラルネットにおいても、中間ユニットの入出力関数はシグモイド関数、出力ユニットのそれは線形関数とした。単語ベクトルは 5 次元で各要素の値は ± 1 の範囲内に制限した。名詞句ベクトル、すなわち composer の出力も必然

boy	-0.65	-0.33	-0.81	0.79	-0.98
man	-0.57	-0.30	-0.78	0.69	-0.99
tall boy	-0.61	-0.78	-0.63	1.14	-1.85
tall man	-0.62	-0.77	-0.62	1.14	-1.83
old man	0.58	-0.38	-1.05	0.23	-0.73
lion	-0.47	-0.28	-0.62	0.41	-0.91
big lion	-1.08	-0.81	-0.55	1.58	-2.23
old lion	0.59	-0.37	-0.97	0.11	-0.60
hammer	-0.26	-0.32	0.09	-0.92	0.33
spoon	-0.32	-0.37	0.15	-0.91	0.34
big hammer	-0.41	-0.30	0.17	-0.91	0.44
big spoon	-0.38	-0.30	0.18	-0.96	0.48
big	-0.51	-0.23	-0.52	-0.10	0.40
small	-0.52	-0.22	-0.52	-0.10	0.40
old	0.01	-1.00	-0.05	-1.00	-1.00
tall	-0.59	-0.37	-0.59	-0.31	-0.04

表 2: 獲得された単語ベクトルと、それによって合成される名詞句ベクトル (一部)

的に 5 次元となるが、要素の値に制限は設けていない。

100 セッションの終了後、学習によって獲得された単語ベクトル、およびそれによって合成される名詞句ベクトルの一部を表 2 に示す。詳しい解析はまだ行っていないが、

- 名詞に tall が付くと第 2、第 4、第 5 要素の値が増大する
- 名詞に old が付くと第 1 要素の符号が反転する
- 同じ名詞と同じ確率で生起する big と small はほぼ同じベクトルとなる
- Man と boy は、以前の形容詞なしの実験ではほぼ同じベクトルであった (図 2) が、共起する形容詞が異なるために今回は差異が大きくなっている

等の特徴を見てとることができる。

4 おわりに

CAN, composer, decomposer という 3 つのニューラルネットワークを協調動作させることで、格要素に 2 語以上から構成される名詞句が来る場合でも Miikkulainen の「拡張されたバックプロパゲーション」を利用することが可能になった。

今後は今回の実験で得られたデータを詳細に解析すると共に、形容詞が付くことによって主辞名詞の格役割が変化するような入力文を与えたときにどのような学習が行われるか調査する予定である。

参考文献

- [1] Miikkulainen, Risto: *Subsymbolic Natural Language Processing*, MIT Press, 1993.
- [2] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: *Learning Internal Representations by Error Propagation*, in Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* Volume 1, pp. 318–362, The MIT Press, Massachusetts, 1986.
- [3] 高橋直人: ニューラルネットを用いた意味表現形式の自動獲得, 電子情報通信学会技術研究報告 NLC 98-28, 1998.
- [4] Takahashi and Motoki: A Subsymbolic Approach for Acquiring Semantic Representations, *Third International Workshop on Computational Semantics*, Tilburg, 1999.