

H-013

系列の区切りを考慮した単純再帰ネットワークモデル A Simple Recurrent Network Model considering Separation of Sequence

竹谷 仁志[†]
Hitoshi Takeya

下村正夫[†]
Masao Shimomura

三宅章吾[†]
Syougo Miyake

阿曾 弘具[†]
Hiroto Aso

1. はじめに

人の高次の脳機能の一つとして言語機能が挙げられる。言語機能についての研究には、その発達の過程に注目したものがあ。Elman は単純再帰ネットワーク (Simple Recurrent Network; 以下 SRN) を用いたモデルを提案し、文中の次単語の予測を繰り返し学習することで、文の系列から文の構造が獲得可能であることを示した [1]。しかし、Elman によるこの学習では文系列の区切りを考慮に入れていない。Elman は後の研究で SRN の学習の際に、外部から SRN の文脈情報をリセットすることで結果が良くなると述べている。外部からの文脈情報のリセットは恣意的であり、自然な文理解とは言えない。そこで本研究では、発話における文の終わりの沈黙の区間や書き言葉ではピリオド等があることに着目し、文系列の区切りの情報を加えた場合に SRN がどのような挙動を取るかを調べた。特に系列の区切りとなる部分に注目し、学習の結果この部分で文脈情報をリセットさせるようになることが期待される。

2. 言語獲得過程のモデル

Elman の SRN を用いた学習では、学習者 (SRN) が周囲でなされる会話 (文系列) を聞き、次に話される単語を予測しているうちに、結果として言語を獲得するものと考えている。使用する単語は 29 種類とし、SRN への入力としては 29 次元の単位ベクトルを用いている。1 の位置が単語の種類を表す。文は SV または SVO の文型をとる名詞と動詞のみからなるとし、語尾変化や句読点等は考慮しない。以上の条件で多数の文を生成し、切れ目なく繋げたものを SRN への入力系列としていた。

SRN は三層パーセプトロンに隠れ層の値を一時保存する文脈層を付加したものである (図 1)。文脈層の付加により、過去の入力 (文脈情報) を反映させた出力が可能となる。入力層と出力層の各ノードを各単語に一対一対応させて入力層と出力層のノード数は 29、隠れ層と文脈層のノード数は 150 とする。

ここで文の生成方法について説明する。29 種類の単語は人を表す名詞や動物を表す名詞、自動詞や他動詞など 12 種類に分類され、これをカテゴリーと呼ぶ。文の生成には、カテゴリーをどのように並べるかが記述されている 16 種類のテンプレートをを用いる。一つのテンプレートをランダムに選択し、さらにテンプレートに記述されているカテゴリーの中から単語をランダムに選択して、一つの文を生成する。

次に、SRN の学習課題について説明する。SRN は入力された単語の次に続く単語を予測することを課題とする。過去の入力によって、同じ入力でも次に続く単語の種類は変わってくるようになるため、予測を正しく行う

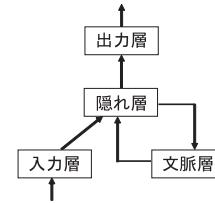


図 1: 単純再帰ネットワーク (SRN)

ためには文脈情報が重要となる。

3. 系列の区切りの考慮

Elman の学習では系列の区切りが明示的に与えられていない。しかし、会話では文と文との区切りの部分で空白の時間があり、区切りを表す情報は会話中に存在しているものと考えられる。そこで本研究では系列の区切りを表す単語 period を加え、文の区切りの情報を SRN に明示的に与えることとする。入力ノード、出力ノードの数を一つ増やし、period を表すノードとする。

本モデルでは文と文は独立しており、文頭の主語が入力となる場合の予測には文脈情報は不要となる。SRN は過去の入力情報を文脈層に蓄積し、出力に影響を及ぼすという性質を持つ。従って、学習後の SRN には period の入力に対して文脈層の値をリセットする効果が現れ、次の主語となる単語が入力されたときに正しい出力が出来るようになることが期待される。文脈情報のリセットの効果が出た場合に、SRN が period を文の区切りとして扱っているとみなすことが出来る。

4. 評価方法

本節ではモデルの評価方法について説明する。

4.1 COSINE による評価

単語 x が入力されたときの出力層の各ノードの出力を x の次に入力される単語の生起確率の予測値と考える。 x は入力系列中に複数回現れるので、予測値の平均を求める。これをさらにカテゴリーごとに平均したものを 12 次元ベクトル O_x とする。本研究では、この O_x と文の生成規則から算出される予測値の理想値 P_x の類似度を表す $COSINE_x$ を $COSINE_x = \frac{O_x \cdot P_x}{\|O_x\| \|P_x\|}$ と定義する。 $COSINE$ が高い値となり、理想の生起確率に近い出力をとった場合に、文の構造を獲得出来たと評価する。

4.2 文脈層の評価

文脈情報は文脈層の各ノードの値を要素とした 150 次元のベクトルで表される。単語 x が入力された時の文脈層の状態を C_x とする。period が入力された時の C_{period} とその平均 \bar{C}_{period} との平均二乗誤差 MSE_{period} を算出し、period 以前の文脈情報によらず一定の値になるかを

[†]東北大学大学院工学研究科

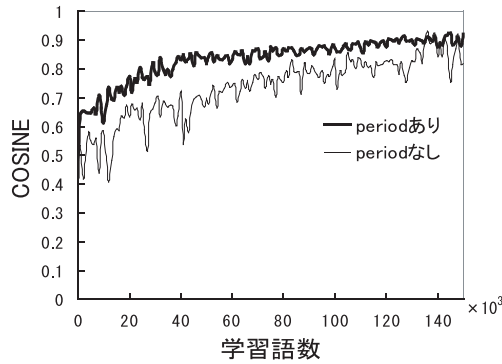


図 2: COSINE の推移

表 1: 平均と period

	$MSE_{\bar{x}}$	MSE_{period}
periodあり	1.08×10^{-2}	5.28×10^{-4}
periodなし	3.76×10^{-3}	1.74×10^{-3}

表 2: 主語と目的語

	MSE_{boy_S}	MSE_{boy_O}
periodあり	1.20×10^{-4}	5.38×10^{-4}
periodなし	4.23×10^{-4}	6.81×10^{-4}

評価する。比較のため、30 個の単語の平均からなる $C_{\bar{x}}$ と $\bar{C}_{\bar{x}}$ との平均二乗誤差 $MSE_{\bar{x}}$ を算出する。 $MSE_{\bar{x}}$ に比べて MSE_{period} の方が小さい値となれば、period が他の単語に比べて文脈層を一定にしていると評価出来る。また、Elman の入力系列についても文末の単語の入力時の文脈層の値を系列の区切りの部分として、平均二乗誤差による評価値を MSE_{period} とした。

さらに、主語と目的語の両方となる単語についてそれぞれの場合の平均二乗誤差 MSE_{x_S} と MSE_{x_O} を算出して比較する。系列の区切りが獲得されている場合、文脈の情報が出力に影響を及ぼす目的語の方が値にばらつきがあらわれ、平均二乗誤差は大きくなると考えられる。

5. 実験による評価

period なし、ありの入力系列について実験を行い、評価する。period なしの学習では 29 個の単語からなる 164670 語の単語系列を入力系列として使用した。period ありの学習では 30 個の単語からなる 150174 語の単語系列を入力系列として使用した。SRN の結合強度、各ノードの閾値の初期値をランダムに与え、BP 法による学習を行った。1000 語学習するごとに COSINE を算出し、学習回数に対する推移を確認した (図 2)。period ありの学習では学習時に早い段階から COSINE が高い値になっており、系列の区切りを表す単語を入力系列に加えることにより学習が速く進むことが確認された。

次に、period なしの学習と period ありの学習のそれぞれについて $MSE_{\bar{x}}$ と MSE_{period} の 100000 回学習時の結果を示す (表 1)。period ありの方が period なしの場合に比べて $MSE_{\bar{x}}$ の値が大きくなっている。これは、同じ単語を入力した場合でも以前に入力された単語の違いによって、文脈層の各ノードの値を大きく変えていることを示す。つまり、period ありの方が過去の文脈の情報を強く出力に反映した学習結果を与えるモデルだと言える。次に MSE_{period} に注目すると、 $MSE_{\bar{x}}$ に比べて小さ

表 3: period なし

	$E[\bar{C}_{\text{boy}}]$	$V[\bar{C}_{\text{boy}}]$
主語	3.20	1.66×10^{-3}
目的語	3.28	1.92×10^{-2}

表 4: period あり

	$E[\bar{C}_{\text{boy}}]$	$V[\bar{C}_{\text{boy}}]$
主語	3.22	1.66×10^{-3}
目的語	3.44	1.92×10^{-2}

い値となっていることがわかる。特に、period ありの方は非常に小さな値となっている。これは period の入力により、それ以前の入力の違いに関わらず文脈層の状態を一定とし、文脈層のリセットの効果が現れていることを示す。つまり、入力系列中に区切りとなる部分があることを SRN が学習できていると評価出来る。

次に主語と目的語の両方となる単語の入力に注目する。今回はその一つとして boy に注目し、 MSE_{boy_S} と MSE_{boy_O} の値を、period ありと period なしのそれぞれについて示す (表 2)。どちらにおいても、主語に比べて目的語の方が平均二乗誤差の値は大きくなっている。ただし、値の比を考えると、period ありの方が大きくなっており、過去の入力情報を反映させて出力を変える傾向が period なしより強くなっているものと考えられる。

また、period ありで C_x の大きさに注目したところ、 \bar{C}_{boy_S} と \bar{C}_{boy_O} とで大きさが異なることが確認された (表 3、表 4)。period ありにおいては $|\bar{C}_{\text{boy}}|$ の平均値に差があり、分散を考慮に入れると主語の場合の boy と目的語の場合の boy とを分類して扱っていると考えられる。目的語の方の分散が大きくなっているのは、過去の入力の違いにより値が変わるためと考えられる。一方 period なしについては平均値の差は小さくなっている。また、分散の方にも大きな差は見られない。これは、period なしでは主語と目的語の違いを SRN が明確に区別できていないことを示唆する結果である。

6. まとめ

系列の区切りの考慮がなかった Elman の学習法に文の区切りを導入した入力系列を考え、実験により評価した。系列の区切りを加えた入力系列による学習では区切りを表す単語が入力されると文脈層の状態が一定となるという文脈層のリセットを SRN がするようになることが確認された。Elman は学習の際に文脈層の値を 3 語ごとに強制的にリセットしつつ学習させることで学習結果がよくなると述べている。本研究で学習初期の段階から評価値が高い値をとったのは文脈層のリセットを SRN 自身がするようになっていたためかもしれない。そのため、文脈層のリセット効果が学習の初期の段階で得られているのかどうかを確認する必要がある。

参考文献

- [1] J. L. Elman. "Finding Structure in Time," Cognitive Science, vol.14, pp.179-211, 1990.
- [2] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett, "Rethinking Innateness," MIT Press, Cambridge, 1996.