

## ニューラルネットを用いた係り受け制約による文生成\*

4 T - 7

吉池紀子 武藤佳恭†

慶応義塾大学政策・メディア研究科‡

## 1 はじめに

本研究は、現代語を用いたいろは歌を作成するための文生成を行うことを目的とする。いろは歌とは、日本語の46文字の仮名を1度ずつ重なりなく使うことによって作られた歌のことである。日本語文節の語群の中からのいろは歌の条件を満たすような文節を選択する操作は組み合わせ最適化問題として捉えることができ、ニューラルネットワークを利用して解を求めることができる [1]。

本論文では [1] によって選択された文節を日本語の文法規則に従って並び替えるための手法を提案する。本手法では、文生成の問題を文節同士の係り受け関係を定める組み合わせ最適化問題として定義した。

他の文生成の手法としては、チョムスキーの提唱した変形文法を用いる方法が一般的に知られている [2]。変形文法とは、句構造文法を用いて文の基本構造を表す構文木（深層構造）を生成し、この深層構造に一連の変形操作を施して表層構造を生成する方法である。この変形文法は、中間状態である句構造を用いるため手続的な生成法ということができ、特に日本語の場合は変形のための文法理論が複雑であるという欠点を持っている [3]。この変形文法を用いる手法に対し、本手法は係り受け制約を用いることにより複雑な文法が必要ないという利点がある。

## 2 係り受けに伴う制約条件

係り受けに伴う制約条件には、以下のものが挙げられる。係り受け制約に関する詳細は [4] 参照。

## 1. 接続可能性：2文節単位でみたときに文法的

\*A Neural Network Approach for Generating Japanese Sentence Based on Dependency Grammar

†Noriko Yoshiike, Yoshiyasu Takefuji

‡Graduate School of Media and Governance Keio University

に接続可能であるものと係り受け対応する。

2. 非交差性：文節間の係り受けは互いに交差しない。
3. 係り先専有性：最後の文節を除いたすべての文節は自分より後方の文節いずれか一つに係る。
4. 卑近接続性：「距離的に近い」文節間ほど係り受けが成立しやすい。

## 3 ニューラルネットワークワークモデル

まず、1. 接続可能性、3. 係り先専有性、4. 卑近接続性を考慮した係り受けペアをつくる際の最適化問題を相互結合型ニューラルネットワークを用いて解く。

ニューラルネットワークで最適化問題を解くために、ニューロン  $V_{ij}$  ( $i, j = 0, \dots, n-1$ ) ( $n$  は文節の数) を用意し、文節  $i$  から文節  $j$  へ係り受けが行われることを  $V_{ij} = 1$ 、行われないことを  $V_{ij} = 0$  と表す。係り受けペアをつくる際の制約条件を考慮したニューロン  $i, j$  の動作式は以下のようになる。

$$\begin{aligned} \frac{dU_{ij}}{dt} = & -A \times (f_1(\sum_{a=0}^{n-1} V_{ai} + \sum_{a=0}^{n-1} V_{ia})) \quad (1) \\ & + f_1(\sum_{a=0}^{n-1} V_{aj} + \sum_{a=0}^{n-1} V_{ja}) - B \times f_2(\sum_{a=0}^{n-1} V_{ia}) \\ & - C \times \sum_{a=0}^{n-1} V_{aj} - D \times 2(1 - d_{h_i h_j}) \times V_{ij} \end{aligned}$$

ただし、ニューロン  $V_{ij}$  は式 (2) で表されるヒステリシス-マッカロックピッツニューロンモデルで、ニューロン  $V_{ij}$  の内部状態  $U_{ij}$  は一次のオイラー法で変化させるものとする。

$$\begin{cases} V_{ij} = 1 & \text{if } U_{ij} \geq 0 \\ V_{ij} = 0 & \text{if } U_{ij} \leq -1 \\ V_{ij} = V_{ij} & \text{otherwise} \end{cases} \quad (2)$$

動作式の第一項目は係り受けのペアが必ず一つ存在するように働く項であり、関数  $f_1(x)$  は式(3)によって定義される。

$$\begin{cases} f_1(x) = -1 & \text{if } x < 1 \\ f_1(x) = 0 & \text{otherwise} \end{cases} \quad (3)$$

第二項目は係り先専有性を満たすためのものあり、関数  $f_2(x)$  は式(4)によって定義される。

$$\begin{cases} f_2(x) = 1 & \text{if } x > 1 \\ f_2(x) = 0 & \text{otherwise} \end{cases} \quad (4)$$

第三項目は卑近接続性による係り受けに伴う距離のコストを示しており、一つの文節へ係り受けが集中しないようにしている。

第四項目は接続可能性を満たすための項である。 $d_{h_i, h_j}$  は文節  $i$  に含まれる単語  $h_i$  と文節  $j$  に含まれる単語  $h_j$  が係り受け可能かどうかを示しており、係り受け可能なときに1、可能でないときに0とする。係り受け可能な文節間の関係は次の3種類の文節関係とする。(連用詞句 → 動詞句), (連体詞句 → 名詞句), (名詞句 → 動詞句)

## 4 文生成

次に、2. 非交差性を満たす語順に並替えるために、文節間の係り受け関係をツリーを使って表す(図1)。ツリーの葉に近い方から順に順番づけをしていくことによって、交差が起るのを防いでいる。

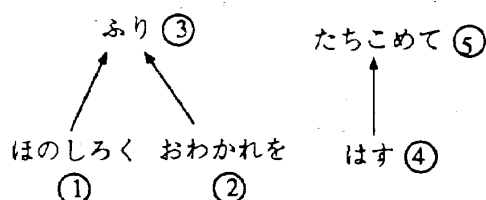


図1 係り受け関係のツリー表記

## 5 シミュレーション結果と考察

シミュレーションでは、動作式の定数、パラメータを  $A = 1, B = 1, D = 1, C = \max(0, 0.5 -$

$0.15 \times (t/50))$  とした。 $C$  をステップ数とともに減少させることによって収束率を高めた。

シミュレーション結果:

ほのしろく おわかれを ふり  
 はす たちこめて  
 やねへ なにも めけようと  
 いえ つむ  
 ゆきらん みせ  
 あさひ そまる

シミュレーションによって作成された文では(いえ → つむ) という意味の通りにくい係り受けが起っていることが指摘できる。今回用いた係り受けの制約に加えて、文節間の係り受け関係に意味が通るかどうかの評価をすることが必要であると考えられる。

## 6 おわりに

シミュレーション結果では人間が手で並替えた文と比較すると係り受けの起る文節間の距離や文節間の係り受けのパターンが似ていた。本手法で文型としては自然な日本語に近い文が生成することが可能であるといえる。しかしながら、現段階では不自然な係り受けが生じる可能性もあり、自然な係り受けや文の意味の評価に関して今後検討したい。

### 参考文献

- [1] 北端美紀, 吉池紀子, 武藤佳恭, 「現代いろは歌を作ろう」 bit1997年8月号, 共立出版
- [2] 長尾真編, 「自然言語処理」, 岩波書店, 1996
- [3] 橋田浩一, 「自然言語の構文解析と文生成の統合」, 情報処理 Vol.33 No.7 p.790-800
- [4] 清水浩行, 佐藤秀樹, 林達也, 「ニューラルネットを利用した日本語係り受け解析 - 係り受け解析の数値問題としての取り扱いについて -」, 電子情報通信学会論文誌 DII Vol.J80-DII No.9 pp.2457-2465 1997年9月
- [5] 武藤佳恭, 「ニューラルコンピューティング」, コロナ社, 1996