自 然 言 語 処 理 108－16
(1995. 7. 21)

ニューラルネットに基づく日本語係り受け解析

清水浩行 * 佐藤 秀樹 † 立岡 章 ‡ 林 達也 *
*名古屋工業大学 電気情報工学科 †日本電装（株） ‡CICC

〒466 名古屋市 昭和区 御器所町 名古屋工業大学 電気情報工学科
TEL: 052-735-5414 FAX: 052-735-5442
E-mail: uniz@maple.elcom.nitech.ac.jp

あらまし　機械翻訳システムへの応用などに代表される自然言語処理技術の一環として、日本語文の文節間係り受け解析がある。我々は係り受け解析を行なうにあたって、4つの構文原理— 非交差性・係り先占有性・卑近接続性・格充足性—に基づいて、文に対する解釈の妥当性に従って増減する評価関数（ポテンシャル・エネルギー）を、各原理の解析に与える影響力の優先度を考慮しつつ定義した。更に評価関数の最小値検索に、最適化問題の求解アルゴリズムとして近年脚光を浴びている相互結合型ニューラルネットワークの分散協調処理能力を、従来の解析システムの逐次検索に代わるものとして利用することを試みた。

キーワード　係り受け解析　日本語解析　ニューラルネットワーク

Dependency Analysis for Japanese Sentence with Mutual Connected Neural Networks

Hiroyuki SHIMIZU * Hideki SATO † Akira TATEOKA ‡ Tatsuya HAYASHI *
*Department of Electrical and Computer Engineering, Nagoya Institute of Technology
†NIPPONDENSO Co., Ltd. ‡CICC

Department of Electrical and Computer Engineering, Nagoya Institute of Technology,
Gokiso, Showa, Nagoya, 466, Japan
PHONE: 052-735-5414 FAX: 052-735-5442
E-mail: uniz@maple.elcom.nitech.ac.jp

Abstract　In this paper we propose to deal with the dependency analysis between *bunsetsu* phrases in Japanese sentences, using the concept of potential energy of interpretation. For each pair of *bunsetsu* phrases in a sentence, if there exists any possible semantic dependency relation between them, binary relation is defined for that pair independently from the rest phrases in the sentence. Each interpretation corresponds to the combination of such binary relations. Potential energy of the interpretation becomes high when structural constraints are violated while it goes down when the constraints are satisfied. Thus the dependency analysis is considered a kind of combinatorial optimization problem and it is solved using mutual connected neural networks. Effectiveness of this approach and some issues are shown.

key words　dependency analysis　Japanese sentence analysis　neural network

1　はじめに

広い意味で一種の最適化問題である係り受け解析を、人間の手に依ることなく機械的に処理するためには、文の解釈の善し悪しを数値の大小で表し、より正しい解釈には高得点を、不適切な解釈には低い点数を与えるような採点方法が必要となる。

言い換えれば、これは文の解釈の妥当性を数値表現できれば、係り受け解析を評価関数の最大もしくは最小化に置き換えることで、（狭義の）**最適化問題**—評価関数を最大化あるいは最小化するような状態を検索する問題—の一つと捉えることができる、ということである。

実際には、文の解釈の妥当性を計るために用いる（妥当に解釈されたあらゆる文において一般的・普遍的に成り立つ）規則、いわゆる**構文原理**にどういったものを用い、またどのように評価関数化するかが重要であり、またこれによってその解析システムの性能が決定されるといっても過言ではないため、往年の係り受け解析の最大の関心事でもあった。

本稿では、その前半部分で係り受け解析を非交差性・係り先占有性・卑近接続性・格充足性といった4つの構文原理を、それらの解釈に対する重要性の序列にも配慮して使って最適化問題化することで、筆者らが構築した解析システムについて紹介し、後半部分では評価関数の最小値検索に、ある種の相互結合型ニューラルネットワークが、そのポテンシャル・エネルギーを最小化するように動作するという性質を利用することを試みた、その結果について述べる。

2　係り受け解析

本研究でいう（日本語の）**係り受け解析**とは、文節単位に区切られた日本語文に対し、2項関係により意味的に成立し得る文節間の係り受け（図1の実線および破線矢印）の中から、文全体として成立するもの（図1の実線矢印）を選び出すことである。

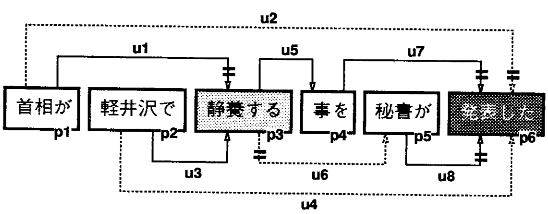


図 1: 文節間係り受け関係の例

2.1　解釈のベクトル表現

係り受け線 i ($i = 1, 2, \cdots, n$; $n \triangleq 2$ 項関係により意味的に成立し得る係り受け関係の総数) を、それが選択されていれば値 "1" を、そうでなければ値 "0" をもつような変数 u_i で表す。このとき、文の係り受けをいかに解釈するかは n 次元の状態ベクトル $\boldsymbol{u} = [u_1, u_2, \cdots, u_n]$ を使って表現できる[1]。例えば、図1の例で妥当な解釈を表すベクトルは、次のようになる。

$$\boldsymbol{u} = \begin{array}{cccccccc} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

2.2　構文原理

本研究では、次の4つの構文原理を評価関数化して係り受け解析に利用した。

1. 妥当な解釈においては、文節間の係り受けは互いに交差しない（**非交差性**）。
2. 妥当な解釈においては、（最後の文節を除き）全ての文節は自分より後方の文節に少なくとも、かつ高々一つ係る。見方を変えれば、任意の文節の係り先は、（それより後方の）ただ一つの文節により占有される、ということである（**係り先の**（1文節による）**占有性**）。
3. 距離的に近い文節間ほど係り受けが成立しやすい（**卑近接続性**）。
4. 一般に文中に複数個存在する用言（動詞、形容詞、形容動詞）等が必要とする**必須格**が満たされるような解釈が、妥当なそれとなる（**格充足性**）。

以下に、実際に解析に用いた評価関数 $E_1 \sim E_4$ を示す ($m \triangleq$ 文節数)。

2.2.1　非交差性の評価関数 E_1

$n \times n$ の相互排他行列 X を用いて、次のように定義する[2]。

$$E_1 = \sum_{i,j=1}^{n} X_{ij} u_i u_j \tag{1}$$

$$X_{ij} = \begin{cases} 1 & \text{係り受け関係 } i \text{ と } j \text{ が交差して} \\ & \text{いる（相互排他的である）場合} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

選択されている係り受け線同士が交差していれば増加。（i 側から見たものと j 側から見たものの2回、カウントするので）交差1つに対して+2。正解時には交差が1つもなくなり、$E_1 = 0$ となって全状態中評価値最低となる。

[1]逆にいえば、2項関係により意味的に成立し得る係り受け関係が n 個、存在するような文では 2^n 通りの係り受け解釈ができるということである。

[2]当然 X は対称行列となる。

2.2.2 係り先占有性の評価関数 E_2

$m \times n$ の係り元行列 Y を用いて、次のように定義する[3]。

$$E_2 = \sum_{k=1}^{m} \left(\sum_{i=1}^{n} Y_{ki} u_i - 1 \right)^2 - 1 \qquad (3)$$

$$Y_{ki} = \begin{cases} 1 & \text{係り受け関係 } i \text{ の係り元が} \\ & \text{文節 (phrase) } k \text{ である場合} \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

各文節についての「出係り受け線の過不足数」の2乗和。

1文節から出ている選択された係り受け線に過不足が生じる毎に2次関数的に増加。正解時には1文節から1本ずつ線が出ている状態となり、$E_2 = 0$ となってこれも全状態中で最低の評価値となる。

2.2.3 卑近接続性の評価関数 E_3

n 次元の文節間距離ベクトル Z を用いて、次のように定義する。

$$E_3 = \sum_{i=1}^{n} Z_i u_i \qquad (5)$$

$$\begin{aligned} Z_i = &\ (\text{係り受け関係 } i \text{ の係り先文節番号}) \\ &- (\text{係り受け関係 } i \text{ の係り元文節番号}) \qquad (6) \\ &\qquad (\leq m-1) \end{aligned}$$

選択されている「係り受け線の長さ」の和。これのみ正解時でも全状態中最低とはならない。

2.2.4 格充足性の評価関数 E_4

$m \times n$ の必須格候補行列 P および m 次元の必須格数ベクトル N を用いて、次のように定義する（図1の例では、必要必須格数を各ノードの濃淡で、必須格候補係り受け線を "=" で表現してある）。

$$E_4 = \sum_{k=1}^{m} \left(\sum_{i=1}^{n} P_{ki} u_i - N_k \right)^2 \qquad (7)$$

$$P_{ki} = \begin{cases} 1 & \text{係り受け関係 } i \text{ が成立することで} \\ & \text{文節 } k \text{ に含まれる用言の必須格が} \\ & \text{一つ埋まる場合} \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

$$\begin{aligned} N_k = &\ \text{文節 } k \text{ に含まれる用言の必要とする必須格数} \\ &(\text{文節 } k \text{ が用言を含まないなら } 0) \qquad (9) \end{aligned}$$

各文節についての「格充足の過不足数」の2乗和。

用言が欲する必須格数に対し、格充足に過不足が生じる毎に2次関数的に増加。これも正解時には全状態中最低 ($E_4 = 0$) となる。

[3] 行列 Y の第 m 行（最終行）は最後の文節に対応するため、必ずオール0となる。そのため式(3)の第1項は常に最低でも評価値 "1" をもつ。それを第2項の "−1" で相殺してある。

以下に、図1の例について各行列・ベクトルを挙げておく。

$$X = \begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \end{array} \begin{array}{c} \begin{array}{cccccccc} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$Y = \begin{array}{c} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{array} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$Z = \begin{bmatrix} 2 & 5 & 1 & 4 & 1 & 2 & 2 & 1 \end{bmatrix}$$

$$P = \begin{array}{c} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{array} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

以上、4つの評価関数に対して、次のようにその和をとることで、4つの原理の成立度の全てを反映するような評価関数が得られる。ここに、a, b, c, d は各原理の解析に対する影響力を表す整数定数であり、これについては次節で詳しく述べる。

$$E = aE_1 + bE_2 + cE_3 + dE_4 \qquad (10)$$

2.3 構文原理の序列—解析貢献度

4つの構文原理の評価関数のうち、卑近接続性に関するものだけは正解時に最小値をとらない。よって、単純にこれらの和をとっただけの関数では、正解時に最小値をとらない場合も起こり得るので、係り受け解析の評価関数としては不適切である。

そこで、これら4つの原理が実際の係り受け解析において果たす役割の大小を、全体の評価関数にも反映させ、それによって 2^n の全ての状態の中で妥当な解釈となるものが、真に最小の評価値を与えるようにする。そのために、従来、経験的に良いものを採用していた各評価関数の係数 a, b, c, d を数式的に決定することを考える。

例えば、卑近接続性の影響を小さくしたければ、その評価関数である E_3 の係数 c の値を小さく、逆に、係り先占有性の影響を大きくしたければ、E_2 の係数 b の値を大きくとればよいわけだが、本研究では、4つの評価関数

がもつ性質に着目して、対象、つまり解析文毎に $a \sim d$ を設定する方法を採る。

今後、これらの係数をその役割から（解析）**貢献度**と呼ぶことにする。それでは、具体的に 4 つの貢献度がどういった形（関数）となるか、その導出も含めて以下に述べる。

2.3.1 方針・基準の設定

貢献度を決めるために、構文原理に関する次の 4 つの点に着目する。

- E_1, E_2, E_4 はそれぞれ正解時に最小となるので、それらの和 "$aE_1 + bE_2 + dE_4$" も必ず正解時に最小となる。従って、貢献度のとり方次第で全評価値が正解時より下回る可能性をもつのは "E_3 が正解のそれより小さい状態" であり、a, b, d は互いに独立に決めても構わない。
- E_1, E_4 が交差や格充足に関わる係り受け線の挙動（選択の有無）以外には無関心であるのに対し、E_2 および E_3 は原理の性質上、全ての係り受け線の挙動に依存する、つまり状態が異なれば（偶然、同じ値をもつものは別として）必ず異なる値をもつ。
- 卑近接続性を表す評価関数 E_3 は、正解にあっても最小値をとらない。
- 他の 3 つの原理が正解であるための必要条件であるのに対し、卑近接続性は必ずしも成り立つとは限らない補助的な条件である。従って、卑近接続性によって他の原理がその成立を妨げられるようなことがあってはならない。

これらの事実を踏まえて、貢献度を定式化していく。

まず、$a \sim d$ の全てに共通して言えることだが、一般にこれらの値を大きくとるほど評価関数の起伏は急峻になる。このことは、後述のニューラルネットを使った最小値検索にとって極小値に嵌まりやすくなるため、あまり好ましくない。従って、これら貢献度の値はできる限り小さくとることが望まれる。

また上述のような理由から、卑近接続性の貢献度である c は大きくとるべきではない。そこで、

$$c = 1 \tag{11}$$

として、これをもって貢献度の基準とすることにする。

このようにして基準が定まったので、システムの状態に評価値がどう依存するかを貢献度に着目して考え、他の貢献度の値を貢献度 c に対して相対的に決定する。

方針としては、不正解状態のうち卑近接続性の評価値 E_3 が正解のそれより小さいもの、つまり、

$$\Delta E_3 \triangleq E_3(\text{正解状態}) - E_3(\text{不正解状態}) > 0 \tag{12}$$

である不正解状態全てについて、いかに他の原理でその影響を殺してやり、トータルとしては正解よりも高い評価値をもたせるようにするかを、式の上で大小関係に着目して考えることになる。

ここで、以下の説明のために E_3 の特殊な場合として、次の D_1, D_2 を定義する。

係り先占有性が成立している ($E_2 = 0$) 状態のうちで、どの文節についても、そこから出ている係り受け線の中で最大の長さをもつものが選択されているような状態の E_3 を D_1、最小の長さをもつものが選択されているような状態の E_3 を D_2 とする。つまり D_1, D_2 は、"$E_2 = 0$" なる状態が与える E_3 のうち最大および最小のものをそれぞれ表している。

$$\begin{aligned} D_1 &\triangleq \max\,(E_3)\Big|_{E_2=0} \\ &= \sum_{k=1}^{m} \left\{ \max_{1\le i\le n} (Y_{ki} Z_i) \right\} \qquad (13) \\ &\left(m - 1 \le D_1 \le \sum_{k=1}^{m-1} k = \frac{(m-1)m}{2} \right) \end{aligned}$$

$$\begin{aligned} D_2 &\triangleq \min\,(E_3)\Big|_{E_2=0} \\ &= \sum_{k=1}^{m} \left\{ \min_{1\le i\le n} (Y_{ki} Z_i) \right\} \qquad (14) \\ &(m - 1 \le D_2 \le D_1) \end{aligned}$$

2.3.2 係り先占有性に視点を置いた状態の場合分け

先に述べたように、卑近接続性についての評価関数のほかに、係り先占有性についてのそれも各状態が固有の値をもつ。そこで係り先占有性にとって大きな意味をもつ "正解時に選択される係り受け線の本数 $(m-1)$" を基準にして、$\Delta E_3 > 0$ なる不正解状態を大きく 2 つのグループに分ける。つまり、

- 選択された係り受け線が、$(m-1)$ 本未満か、m 本以上の状態
- 選択された係り受け線が、正解時と同じ $(m-1)$ 本である状態

にである。

ここで、これらの状態について、それぞれ考える前に ΔE_3 として考えられる最大のもの

$$\begin{aligned} \max\,(\Delta E_3) = \max\,(E_3(\text{正解状態})) \\ - \min\,(E_3(\text{不正解状態}))\ (\ge 0) \quad (15) \end{aligned}$$

が、どういった形をとるか求めてみる。右辺第 1 項は物理的には、"正解として考えられる状態が与える E_3 のうちで、最大のもの" を意味し、これは D_1 に他ならない。また、第 2 項は上で分けた場合毎に異なる。

2.3.3 選択された係り受け線が、$(m-1)$ 本未満か、m 本以上の状態について

選択された係り受け線が $(m-1)$ 本でないということは、「1つの文節から1本の係り受け線」という係り先占有性は絶対に満たされていないことを意味する。

このことから、選択された係り受け線が $(m-1)$ 本未満の状態では、その不足数を l (lack) 本とすれば、最低でも l は正解時より係り先占有性の評価関数 E_2 に関しては高くなる。同様に選択された係り受け線が m 本以上の状態では、超過数を e (excess) 本とすれば、少なくとも e は高くなる。

任意の解析文に対しての臨界条件を求めるためには、正解状態との E_2 についての評価値の差

$$\Delta E_2 \triangleq E_2(\text{不正解状態}) - E_2(\text{正解状態}) \ (>0) \tag{16}$$

が最小となる不正解状態が、$\max(\Delta E_3)$ を与える場合について考えるのが妥当である。

ここで、選択された係り受け線が $(m-1)$ 本未満なら $\min(\Delta E_2) = l$、m 本以上なら $\min(\Delta E_2) = e$ となるので、上述の場合にシステム的に不正解、つまり正解状態よりも高い総評価値を $\max(\Delta E_3)$ を与える不正解状態がとるためには、貢献度 b は次の2つの条件をそれぞれ満足せねばならない[4]。

i) 選択された係り受け線が $(m-1)$ 本未満の場合：

$$bl > c\max(\Delta E_3) \tag{17}$$

ii) 選択された係り受け線が m 本以上の場合：

$$be > c\max(\Delta E_3) \tag{18}$$

それでは、上のそれぞれの場合について詳しく検証する。

i) $(m-1)$ 本未満の場合：

この時、選択された係り受け線の本数は $\{(m-1)-l\}$ と表せ、(15) 式右辺第2項は、これらの係り受け線全てが長さ $\min(Z_i)$ である[5] ような不正解状態によって与えられる。つまり、

$$\min\ (E_3(\text{不正解状態})) = \{(m-1)-l\}\min\ (Z_i) \tag{19}$$

となる。従って、式 (17) は次のようになる。

$$\begin{aligned} bl &> c\,[D_1 - \{(m-1)-l\}\min(Z_i)] \\ \Leftrightarrow \quad & l\,\{b - c\min(Z_i)\} \\ &> c\,\{D_1 - (m-1)\min(Z_i)\}\ (\geq 0) \end{aligned} \tag{20}$$

[4]臨界条件を求めるのであるから、非交差性と格充足性は成り立っているものとすることに注意。

[5]長さが同じということは、これらの係り受け線は必然的に異なる文節から生じていることになる。

この式で、臨界状態となるのは $l=1$ の時で、結局、選択された係り受け線が $(m-1)$ 本未満の状態に関しては、次式が成り立てばよいことがわかる。

$$b > c\,\{D_1 - (m-2)\min(Z_i)\} \tag{21}$$

ここで $\min(Z_i)$ は、後ろから2番目の文節 p_{m-1} は最後の文節 p_m にしか係ることができないことから、文に依らず $\min(Z_i) = 1$ である。

ii) m 本以上の場合：

この時、選択された係り受け線の本数は $\{(m-1)+e\}$ と表せ、(15) 式右辺第2項は、これらの係り受け線全てが“できる限り短い長さ”をもっているような不正解状態によって与えられる。従って $\min(Z_i) = 1$ として、k 番目に短い係り受け線の長さを単調増加関数 $g(x)$ を用いて $\min(Z_i) + g(k)$ $(0 \leq g(k) \leq m-2)$ と表すことにすれば、(15) 式の右辺第2項は次のようになる。

$$\min\ (E_3(\text{不正解状態})) = \{(m-1)+e\}\min(Z_i) + \alpha(e) \tag{22}$$

$$\alpha(e) \triangleq \sum_{k=1}^{(m-1)+e} g(k)$$

よって、式 (18) は α も単調増加関数であることから次のようになる。

$$\begin{aligned} be &> c\,[D_1 - \{(m-1)+e\}\min(Z_i)] \\ &> c\,[D_1 - (\{(m-1)+e\}\min(Z_i) + \alpha)] \\ \Leftrightarrow \quad & e\,\{b + c\min(Z_i)\} > c\,\{D_1 - (m-1)\min(Z_i)\} \\ &> c\,\{D_1 - ((m-1)\min(Z_i) + \alpha)\} \end{aligned} \tag{23}$$

この式で、臨界状態となるのは $e=1$ の時で、結局、選択された係り受け線が m 本以上の状態に関しては、

$$b > c\,\{D_1 - m\ \min(Z_i)\} \tag{24}$$

が成り立てばよいことがわかるが、これは次のように (21) 式に較べて甘い条件であるので無効となる。

$$\begin{aligned} b &> c\,\{D_1 - (m-2)\min(Z_i)\} \\ &> \underline{c\,\{D_1 - m\ \min(Z_i)\}} \end{aligned} \tag{25}$$

2.3.4 選択された係り受け線が、正解時と同じ $(m-1)$ 本である状態について

これらの状態を更に、係り先占有性が成り立っているか、否かで分けて考える。

i) 係り先占有性不成立時：

選択された係り受け線が $(m-1)$ 本であるのに係り先占有性が成立していないということは、1本も係り受け

線が出ていない（どこにも係っていない）文節が幾つかあり、その分、線が複数本出ている文節が存在するということである。

従って、1本も係り受け線が出ていない文節の数を s (separated) とすれば、$\min(\Delta E_2) = 2s$ となる[6]。また、先の選択された係り受け線が m 本以上の場合と同様、この時もどの選択された係り受け線の長さも $\min(Z_i)$ より短いことはなく、

$$\begin{aligned} &\min\,(E_3(\text{不正解状態})) \\ &\qquad = (m-1)\min(Z_i) + \beta(s) \end{aligned} \tag{26}$$

(β：単調増加関数) と表せる。以上から、

$$2bs > c\,[D_1 - \{(m-1)\min(Z_i) + \beta\}] \tag{27}$$

とならねばならない[7]。しかし、この式も次のように (21) 式が成り立てば自ずと満たされる。

$$\begin{aligned} \underline{2bs} > b > c\,\{D_1 - (m-2)\min(Z_i)\} \\ > \underline{c\,[D_1 - \{(m-1)\min(Z_i) + \beta\}]} \end{aligned} \tag{28}$$

ii) 係り先占有性成立時：

$\max(\Delta E_3)$ は、係り先占有性の成立している不正解状態に対しては "$D_1 - D_2$" となり、この時、D_2 を与える状態が係り先占有性の成立にも拘らず不正解であるというのだから、交差が少なくとも1つは存在するか、格充足性が完全には満たされていないはずである。従って、次の2つの臨界条件がそれぞれ成り立つ必要がある。

$$2a > c\,(D_1 - D_2) \tag{29}$$

$$d > c\,(D_1 - D_2) \tag{30}$$

2.3.5 結論

以上の不等式群から、貢献度を整数とするなら、

$$\left\{ \begin{array}{l} b = c\,\{D_1 - (m-2)\} + 1 \\ a = \{c\,(D_1 - D_2)\ \mathrm{div}\ 2\} + 1 \\ d = c\,(D_1 - D_2) + 1 \end{array} \right. \tag{31}$$

$(b > d \geq a \geq c = 1)$ と表せる[8]。

図1の例で具体的に貢献度を計算してみれば、次のようになる。

$$b = 11,\ a = 4,\ d = 8$$
$$(\because\ D_1 = 14,\ D_2 = 7;\ c = 1)$$

[6]一つ、係り受け線の出ていない文節があれば、同時に線が最低でも2本は出ている文節が存在するので、2が掛かる。

[7]臨界条件を求めるのであるから、非交差性と格充足性はここでも成立しているものと考える。

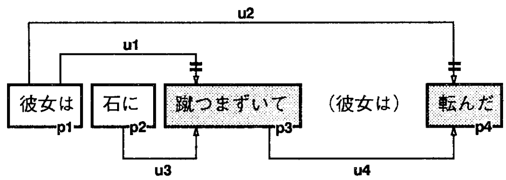
[8]div は整数除算の商を返す二項演算子。

2.4 逐次検索による解析結果

以上で得られた評価関数を用いて、実際にいくつかの例文について逐次検索により解析— 評価関数の最小値検索—を行なった。その結果、本システムが完全な単文・複文に対しては、評価値最小状態として妥当な解釈を得られる、係り受け解析システムと成り得ることが実証できた。

一方、本システムでは正確な解析が不可能な文のほとんどは、以下に示す重文か、自明格の省略された文としては不完全なものの二種類に大別されることもわかった。

2.4.1 重文における係り先共有性



上図のような、いわゆる**重文**では重複する語の省略により、一つの格が複数の用言に係る（係り先が複数文節により共有される）ことが、ごく当たり前にある。しかし、今回用いた構文原理では係り先占有性により一つの文節は、それより後方の一文節にしか係ることを許されないので、非交差性や格充足性が影響しなければ、卑近接続性から一番近くに接続している係り受け線のみが選択される結果となっている。当然このとき、共有されるべき語が必須格であれば格充足性が満たされず、他の係り受け線の選択の誤りのきっかけにもなりかねない。

2.4.2 自明格の省略に伴う格充足の異常

用言の必須格として、主格は必ず挙げられるが、その反面、主格はそれ以前の文の流れから "言わなくてもわかる" 暗黙の了解事項として、省略されることも多い（この暗黙のうちに省略され得る主格を**自明（な）格**と呼ぶ）。

自明格の省略された文では、正しい解釈でも用言の必須格が全て満たされることはなく、格充足性の評価関数 E_4 は最小値をとらない。

実際の解析では、省略された自明格の分を他の目的格候補が埋めようとして、間違った係り受け線が選択されることが多い。例えば必須格数：主格1, 目的格2の用言に対し、目的格候補が3つ以上ある場合、目的格のみで格充足が完了される、といった具合にである。

以上が、本研究の前半部にあたる『係り受け解析システムの構築』に関する報告である。

本稿の残りの部分では、このシステムに対して評価関数の最小値検索のためにニューラルネットを導入する試みについて、実際に執り行った実験の内容を紹介する。

3 ニューラルネットを用いた最小値検索

まず、最適化問題の求解アルゴリズムとしてのニューラルネットについて説明する。

3.1 Hopfield のネットワーク

ネットワーク内の各ユニット u_i が、

$$I_i = \sum_j w_{ji} u_j - \theta_i \tag{32}$$

$$\begin{cases} w_{ji} : u_j \text{から } u_i \text{への結合の強さ（重み）} \\ \theta_i : u_i \text{の閾値} \end{cases}$$

$$u_i = f(I_i) \tag{33}$$

$$\text{出力関数 } f(x) = u(x) : \text{単位ステップ関数} \tag{34}$$

に従って**非同期的**に状態遷移を行なう相互結合型ニューラルネットワークが、互いに等しい強さで結合—**対称結合**—し ($w_{ij} = w_{ji}$)、自身に対して直接、フィードバック結合をもたない ($w_{ii} = 0$) ならば、ネットワーク全体は次式で与えられる**エネルギー**と呼ばれるネットワークの状態の関数 E を極小化するように動作する。

$$E(\boldsymbol{u}) = -\frac{1}{2} \sum_{i,j} w_{ij} u_i u_j + \sum_i \theta_i u_i \tag{35}$$

この **Hopfield のネットワーク**と呼ばれるニューラルネットは、常に (35) 式が減少するように状態遷移するため、ネットワークはエネルギーの最小状態ではなく極小状態に収束する可能性がある。これを防ぐために確率を導入したのが、次に挙げる BNN である。

3.1.1 BNN と SA

ニューロンの出力関数として次式を用いたのが、**ボルツマン・ニューラルネットワーク (Boltzmann Neural Network; BNN)** と呼ばれるネットワークである。その他の条件等は Hopfield のネットワークとかわりない。

$$\text{遷移確率 } p(u_i = 1) = \frac{1}{1 + \exp(-I_i/T)} \tag{36}$$

この式の重要性は、たとえ入力の線形和が小さくても、ユニットに 1 となる可能性を残しているところにある。これが、Hopfield のネットワークにはなかったエネルギー増加の方向への状態遷移を可能にし、極小値からの脱出を実現した。

ここで、T は**温度**と呼ばれる正の定数で、$T \to 0$ の極限においては式 (36) はステップ関数となり、Hopfield のネットワークと同じに、つまり決定論的な動作をするようになる。このとき、定常状態においてネットワークは、エネルギーの高い状態をとることはほとんどなく、非常に高い確率でエネルギーの低い状態へと落ち着く。

従って、はじめのうちは極小値からの脱出を図るため高温でネットワークを作動させ、しばらくして最小解付近に状態が近付いたら低温へと冷却してやれば、最小解に収束することが期待できる。このように、高温から低温へ、確率論的から決定論的へと系を変化させることで最適解を得ようとする手法は、**シミュレーティッド・アニーリング (Simulated Annealing; SA, 模擬徐冷)** と呼ばれ、実際にはどの程度の速さで冷却してゆくか—**クーリング・スケジュール**—が大きな問題となる。

さて、以上のことから (35) 式に評価関数を対応させ、適当なアニーリングを行なうことができれば、BNN を用いて最適化問題を解くことができることがわかる。

3.2 係り受け解析用のネットワークの設計

ニューラルネットを使った係り受け解析では、各係り受け線にユニットを一つずつ割り当てる。

評価関数 ((10) 式) とネットワークのエネルギー式 (35) との対応をとれば、ネットワークの重み・閾値を与えるものとして次の 2 式が得られるので、これらに従って実際に係り受け解析用のネットワークを設計すればよい。

$$w_{ij} = -2 \left\{ aX_{ij} + \sum_k (bY_{ki}Y_{kj} + dP_{ki}P_{kj}) \right\} \tag{37}$$

$$\theta_i = cZ_i - b - d \sum_k \{P_{ki}(2N_k - 1)\} \tag{38}$$

3.3 実験

本研究ではクーリング・スケジュールには、一般によく用いられる次の関数を用いた (T_0 = 初期温度)。ここに ΔT は**インターバル（下げ幅）**と呼ばれる定数で、図 2 に示されるようにこの値が大きいほどゆったりとした冷却に、小さいほど急冷となる。つまり、冷却スピードをこれで制御することができる。

$$T(t) = \frac{T_0}{1 + t/\Delta T} \tag{39}$$

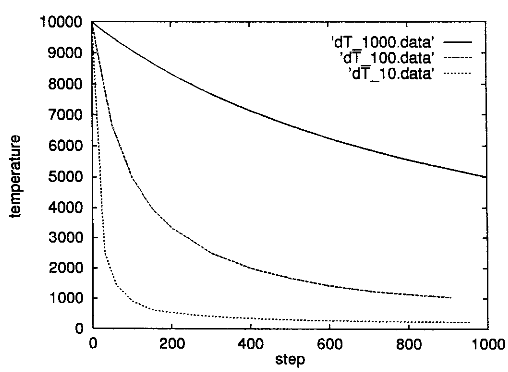


図 2: クーリング・スケジュール

初期温度は小さくとれば、無駄な—エネルギーを増加させる方向への—状態遷移は生じ難くなるが、極小値からの脱出が困難になるため、より一層ゆっくりとアニーリングを行なわねばならなくなる。また、エネルギー関数の起伏に較べて初期温度が極端に低いと、最初に嵌まった極小値から永久に抜け出せなくなるので、初期温度には自ずと小さくとれる限界というものが存在すると考えられる。

以上の観点から、ネットワークを最小値に収束させ得る限界の初期温度、および初期温度に対してインターバルをどこまで小さくとることができるか（どれだけ速く冷却できるか）の限界を実験によって見極めることで、従来、実験による経験から決定していた初期温度、インターバル値について適切な選び方をするための何らかの法則性を見い出そうと試みた。

3.4 実験結果—初期温度に対する冷却スピードの限界

実際に、いくつかの例文に対して構成したネットワークでシミュレートを行なった結果、初期温度やインターバルを上手く調整することで（逐次検索によって事前に得られていた）真の最小値へと系を収束させることができることがわかった。

図 3 は、最小解を得るのに初期温度に対して、インターバルをどこまで小さくとれるか、その限界値をある例文について実際に計測した様子である。

この図から初期温度を小さくとるほど、よりゆったりとした冷却が必要になることが確認できる。また、初期温度が一定値以上でなければならないこともわかる。

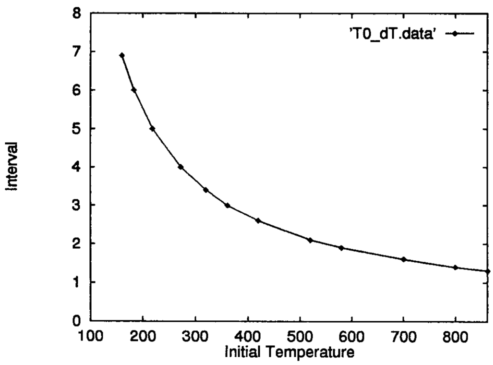


図 3: 初期温度に対するインターバルの限界値

以上のように実験から予想通りの結果が得られたが、初期温度やインターバルの調整により系を最小値に収束させることができたのは、逐次検索により真の最小値が既知であったからであり、本来の最小値が未知の問題に対する初期温度・インターバルの適切な選び方については、残念ながら確固たる解答は得られなかった。

4 おわりに

本稿の前半部分では、いかに係り受け解析を評価関数の最小化という数学的な最適化問題として整備するかについて論じた。

特に貢献度を実験による経験からではなく、数式的に決定する方法は今までにあまり類の見られない新たな試みであり、係り受け解析に限らず、一般の最適化問題にも十分応用が効くものと考えられる。そういった意味では、汎用性の高い手法を提示できたのではないかと思う。

また、本システムの限界を明らかにすることで、重文や自明格の省略された文を処理するためには、より強力な解析システムが必要となることを示唆することもできた。

後半部では、前半部で得られた評価関数の最小値検索にニューラルネットを利用する試みについて述べた。これにより、初期温度やインターバル値のとり方には、ある種の制限があることを明示した。

他方、これらのとり方について文節数やユニット数との間に、何らかの依存関係が発見できることを期待したのだが、十分な結果を得るには至らなかった。これについては（今回は $\boldsymbol{u} = [0 \cdots 0] (= 0_{10})$ としていた）初期状態の選び方と共に、今後も研究を続けていく予定である。

5 参考文献

[1] 野村 浩郷: “自然言語処理の基礎技術”, 電子情報通信学会 (1988)

[2] 合原 一幸: “ニューラルコンピュータ　脳と神経に学ぶ”, 東京電機大学出版局 (1988)

[3] 麻生 英樹: “ニューラルネットワーク情報処理 —コネクショニズム入門、あるいは柔らかな記号に向けて”, 産業図書株式会社 (1988)

[4] 中野 馨 他: “ニューロコンピュータの基礎”, コロナ社 (1990)

[5] 高橋 直人, 板橋 秀一: “ニューラルネットワークを用いた日本語解析の試み”, 情報処理学会論文誌 Vol. 32, No. 10, pp. 1330–1337 (1991)

[6] 立岡 章, 鈴木 勝男, 中村 澄江, 林 達也: “格充足原理に基づく係受け解析手法 —構文原理とそのアルゴリズム—”, 情報処理学会 NL研 Vol. 92, No. 2, pp. 9–18 (1992)