社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2004－NL－162 (29)
2004／7／16

相対的な係りやすさを考慮した日本語係り受け解析モデル

工藤 拓[1]　　松本 裕治

奈良先端科学技術大学院大学 情報科学研究科

〒630-0192 奈良県生駒市高山町 8916-5

{taku-ku,matsu}@is.naist.jp

本稿では, 相対的な係りやすさを考慮する新しい係り受け解析モデルを提案する. 従来の統計的係り受け解析手法の多くは, 元の問題を着目する二文節が係るか係らないかという二値分類問題に帰着させ, 任意の機械学習アルゴリズムを適用していた. しかし, 二文節のみが与えられた状態で係るか係らないかの分別を行うことは一般に困難な場合が多い. 係り受け解析は, 候補集合から係り先を1つ選択するタスクであるため, 二値分類よりは, 候補間での係りやすさの相対的な大小関係を比較するほうがタスクの性質をうまく反映している. 本稿で提案する「相対モデル」は, 係りやすさの相対的な「差」をモデル化し学習することが可能である. 京大コーパスを用いて実験を行った結果, 従来法と比較して学習効率で改善されるとともに, 高い正解率 (91.37%) を示した.

Japanese Dependency Parsing using Relative Preference of Dependency

Taku Kudo　　Yuji Matsumoto

Graduate School of Information Science, Nara Institute of Science and Technology

{taku-ku,matsu}@is.naist.jp

This paper presents a new statistical Japanese dependency parser which models a relative preference of dependency. Most conventional Japanese dependency parsers are based on binary classifications where all possible pairs of segments are classified into positive (dependent) or negative (non-dependent) examples. However, such methods are not suitable for dependency parsing, since the goal of this task is not to *classify* a pair of segments into two classes, but to *select* the most likely modifiee out of all candidates. The proposed method is based on this observation and models how likely a pair of segments have a dependency relation in comparison with other pairs. Experiments using the Kyoto University Corpus show that the method outperforms previous systems as well as improves the training efficiency.

1 はじめに

係り受け解析は言語処理の基本技術として認識されており, これまで多くの研究が行われてきた. 初期の研究では, 二文節間の係りやすさを決定するルールを人手で作成していたが, 網羅性や一貫性という面で問題が多い. 近年では, 構文情報が付与された大規模コーパスが利用可能になったことで, 機械学習アルゴリズムを用いた統計的な構文解析技術が提案されるようになった.

従来の統計的日本語解析の多くは, 他の係り先候補とは独立に算出される絶対的な係りやすさに基づきモデル化されていた. 絶対的な係りやすさは, 候補となる二文節を「係る」か「係らないか」という二値分類問題に帰着することで算出できる. 二値分類器との親和性の高さから, 決定木, 最大エントロピー法, SVM といったな機械学習手法がこれまで適用されてきた [12, 14, 13, 7]. 本稿では, このような絶対的な係りやすさに基づく手法を「**絶対モデル**」と呼ぶ.

しかしながら, 二文節のみが与えらえた状態で, 係るか係らないかを弁別することは困難な場合が多い. 一見正しそうな係り関係であっても, 別の文脈では係

[1] 平成16年4月より NTT コミュニケーション科学基礎研究所 (taku@cslab.kecl.ntt.co.jp)

り関係とならない事例が少なからず存在する．このような事例は，二値分類学習そのものを困難にするばかりでなく，学習後に得られるモデルの信頼性を下げる可能性がある．

一方，日本語の係り受け解析は，係り先候補から正解の係り先を1つだけ選ぶタスクである．そのため，絶対的な係りやすさに基づき係り先を決定するよりは，候補間での係りやすさの相対的な大小関係を比較するほうがタスクの性質をうまく反映している．本稿では，上記の考えに基づき，係りやすさの相対的な「差」をモデル化し学習する係り受け解析手法 **「相対モデル」** を提案する．

相対モデルは，優先度学習 [5] と呼ばれる機械学習手法の日本語係り受け解析への自然な適用となっている．優先度学習は情報検索結果のリランキング [5, 6]，照応解析 [19, 11]，英語の係り受け [11] 等に用いられている．

本稿の構成は以下の通りである．2章で従来法と提案手法の違いを述べ，3章で具体的な学習アルゴリズムについて述べる．4章で絶対モデル以外の従来手法に触れ，本手法との関連性や違いについて言及する．さらに5章で京大コーパスを用いた評価実験を提示し，最後に6章で本稿をまとめる．

2 統計的日本語係り受け解析

本章では，日本語係り受け解析の基本的な性質と，表記法について説明する．

日本語の文に対し，その文節列を $B = \{b_1, b_2, \ldots, b_m\}$，係り受けパターン列 $D = \{d_1, d_2, \ldots, d_m\}$ と定義する．ただし，d_i は，文節 b_i の係り先文節番号を示す．例えば，文節 b_i が文節 b_j に係る場合，$d_i = j$ となる．また，b_i が b_j に係る時，$b_i \rightarrow b_j$ と表記する．これ以降，D は以下の制約を満たすものと仮定する．

(1) 文末を除き，各文節はその文節の後方側に必ず一つの係り先を持つ．

(2) 係り受け関係は交差しない．

制約 (1) より，文末の文節 b_m には係り先が存在しない．そのため，$d_m = -1$ と便宜的に定義しておく．

統計的係り受け解析は，M 個の正解事例 $S = \{\langle B_1, D_1, \rangle, \ldots, \langle B_M, D_M, \rangle\}$ を用い，入力文節列 B から係り受けパターン列 D への写像 $f : \mathcal{B} \rightarrow \mathcal{D}$ を導出するタスクと定式化される．

統計的日本語係り受け解析の多くは，個々の文節の係り先を独立に求める部分問題を扱うことが多い．本稿でもそのような定式化を行う．この時，文節 b_i とその係り先文節番号 d_i のペア $T = \{\langle b_1, d_1 \rangle, \ldots, \langle b_L, d_L \rangle\}$ が正解事例となる．(ただし $L = (\sum_{k=1}^{M} |B_k|)$)．

また，文節 b_i の係り先の候補は，制約 (1) より，b_i の後方にある全文節となる．b_i の係り先候補集合を $C_i = \{b_{i+1}, \ldots, b_m\}$ と表記する．

さらに，二文節を特徴付ける言語的素性ベクトルを $\Phi(\langle b_i, b_j \rangle) \in \mathbb{R}^n$ と表記する．一般には，各文節の品詞や語彙といった情報，二文節の周辺のコンテキスト，あるいはそれらの組み合わせが素性ベクトルとして表現される．

2.1 絶対モデル (従来法)

まず，従来法の絶対モデルについて説明する．絶対モデルでは，候補となる二文節 $\langle b_i, b_j \rangle$ が係り受け関係にあるか (正例) ないか (負例) の二値分類問題を考える．具体的には，言語的素性ベクトル $\Phi(\langle b_i, b_j \rangle) \in \mathbb{R}^n$ の正例負例 $\{+1, -1\}$ への写像 $f : \mathbb{R}^n \rightarrow \{+1, -1\}$ を導出する．既存の二値分類器との親和性が高いことから，これまで決定木，最大エントロピー法，SVM などが絶対モデルに適用されている [12, 14, 7]．

特に，最大エントロピー法，パーセプトロン，SVM といった線形分類器を適用する場合は，以下のような戦略のもと，係るか係らないかを識別する分離平面 $\mathbf{w} \in \mathbb{R}^n$ を導出する．

学習戦略 1 絶対モデル
全文節 b_i と，その候補集合 C_i について，以下の制約を満たすようなベクトル $\mathbf{w} \in \mathbb{R}^n$ を導出せよ．

$$\forall i, \ \forall c \in C_i$$
$$\mathbf{w} \cdot \Phi(\langle b_i, c \rangle) > 0 \ \textit{if} \ c = b_{d_i}$$
$$\mathbf{w} \cdot \Phi(\langle b_i, c \rangle) < 0 \ \textit{if} \ c \neq b_{d_i}$$

b_i が b_j に係るかどうかの判定は，$\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ の符号 $sgn(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))$ で与えられる．また，係りやすさの度合は $\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ で近似できる．文節 b_i の係り先 $\hat{c}_i$ は以下で近似的に与えられる．

$$\hat{c}_i = \underset{c \in C_i}{argmax} \, \mathbf{w} \cdot \Phi(\langle b_i, c \rangle) \qquad (1)$$

2.2 絶対モデルの問題点

絶対モデルは，候補二文節に対する二値分類として定式化された．果たして，このような二値分類は可能なのであろうか．

図1に，絶対モデルでは学習が困難な事例を示す．絶対モデルを適用すると，係り関係 { 太郎は → 探している } は，文1では正例となり，文2では負例とな

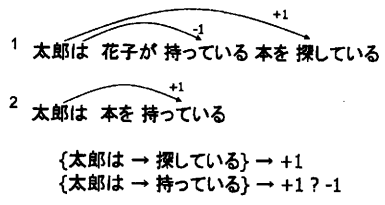

図 1: 絶対モデルで学習困難な例

り, 互いに対立する事例となってしまう. 一般に, 二文節 $\langle b_i, b_j \rangle$ のみが与えらえた状態で, 係るか係らないかを弁別することは困難な場合が多い. 図 1 の例で示すように, 一見正しそうな係り関係であっても, 別の文脈では係り関係とならない事例が少なからず存在する. このような対立する事例は, 学習そのものを困難にするばかりでなく, 学習後に得られるモデルの信頼性を下げる可能性がある.

絶対モデルは, 学習時のみならず, 解析時にも矛盾する事例を生む可能性を持つ. 絶対モデルでは, 係るかどうかの判定は $sgn(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))$ で与えられる. しかし, 解析時に符号が正となるものが複数あったり, 逆に正となるものが一つも存在しなかった場合どのような基準で係り先を決定すればよいのだろうか? 近似的に式 (1) で係り先は決定できるとはいえ, 二値分類が最初の目的であったために, $(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))$ の大きさが係りやすさの度合を表現しているとは必ずしも言えない.

係り受け解析を二値分類で解くために, 我々は, 係り関係そのものを素性として用いる「動的素性」を提案している [16]. 文 1 の場合「持っている」という文節が「花子が」に既に修飾されているという情報が素性として使われる. ただし, 動的素性が絶対モデルにおける問題を全て解決しているとは言えない.

2.3 相対モデル (提案法)

絶対モデルは, 他の係り関係とは独立に係りやすさを算出していた. 一方, 日本語の係り受け解析は, 文節 b_i の係り先候補 $C_i = \{b_{i+1}, \ldots, b_m\}$ から, 正解の係り先を 1 つだけ選ぶタスクである. そのため, 絶対的な係りやすさに基づき係り先を決定するよりは, 候補間での係りやすさの相対的な大小関係を比較するほうがタスクの性質をうまく反映している.

図 2 に絶対モデルと相対モデルの違いを図示する. 与えられた 2 文のそれぞれの文節のペア (A〜F) が, 素性抽出関数 $\Phi(\cdot)$ により素性空間上に配置されたとする. 絶対モデルは, 係るか係らないかの二値分類を基本としており, それらの境界を求めるような学習が行われる. しかし, クラスラベルの対立 (ペア B) や, 例外的な配置 (ペア E) を考えると, その境界 (特に線形境界) を引くのは困難である.

一方, 相対モデルでは, 文節 b_i の係り先候補 $\{b_{i+1}, \ldots, b_m\}$ それぞれに対し, 係りやすさの大小関係が正しく保存されるような射影ベクトル $\mathbf{w}$ を求める. 図 2(右) に具体例を示す. 各文節ペアは $\mathbf{w}$ を用い数直線上に射影される. この数直線は係りやすさの大小関係を示す. この時, 候補間での係りやすさの相対的大小関係が数直線上に保存されていることに注意されたい.

相対モデルの学習は次のように定式化される.

学習戦略 2 相対モデル
全文節 b_i と, その候補集合 C_i について, 以下の制約を満たすようなベクトル $\mathbf{w} \in \mathbb{R}^n$ を導出せよ.

$$\forall\, i,\ \ \forall\, c \in C_i \setminus b_{d_i}$$
$$\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle) > \mathbf{w} \cdot \Phi(\langle b_i, c \rangle)$$

戦略 1 と 戦略 2 の違いに注意されたい. 戦略 1 では, 各 2 文節を正負の二値分類していた. 戦略 2 では, 正解の文節ペア $\langle b_i, b_{d_i} \rangle$ の射影後の値 $\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle)$ が, 他のどの候補の値 $\mathbf{w} \cdot \Phi(\langle b_i, c \rangle)$, $c \in C_i \setminus b_{d_i}$ よりも大きくなるような制約となっている. 係りやすさの相対的な大小関係が重要視されるため, 2.2 節で示したような対立事例の問題は起きにくい.

文節 b_i の係り先 $\hat{c}_i$ は以下で与えられる.

$$\hat{c}_i = \underset{c \in C_i}{argmax}\, \mathbf{w} \cdot \Phi(\langle b_i, c \rangle) \qquad (2)$$

文節 $\langle b_i, b_j \rangle$ の係りやすさの度合は $\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ で与えられる. これは数直線上の値に他ならない.

3 最大エントロピー法による定式化

戦略 2 で示したような学習は, 一般に優先度学習 [5] と呼ばれ, これまで RankBoost[3], SVM の優先度学習への拡張 [6, 2, 10], トーナメントモデル [19] といった手法が提案されている. 本稿では, 学習の効率性を考え, 最大エントロピー法を用いて定式化する.

3.1 定式化

最大エントロピー法による定式化では, 文節 b_i の係り先候補集合 C_i が与えられた時, b_i が $b_j (\in C_i)$ に係る条件付き確率 $p(b_i \rightarrow b_j | C_i)$ を考える.

$$p(b_i \rightarrow b_j | C_i) = \frac{\exp(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))}{\sum_{c \in C_i} \exp(\mathbf{w} \cdot \Phi(\langle b_i, c \rangle))}$$

比較のために, 絶対モデルに最大エントロピー法を適用した場合を以下に示す. これは, 文献 [14] に用い

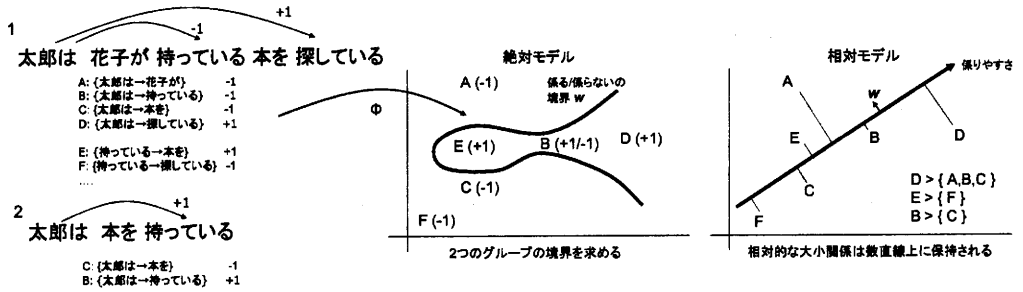

図 2: 絶対モデルと相対モデル

られた手法と同一である. 絶対モデルでは, 文節ペア $\langle b_i, b_j \rangle$ が与えらえた時, それが係る $y = +1$ か係らないか $y = -1$ の条件付き確率を求める.

$$p(b_i \to b_j | \langle b_i, b_j \rangle) = \frac{\exp(\mathbf{w} \cdot \Phi(+1, \langle b_i, b_j \rangle))}{\sum_{y \in \{+1,-1\}} \exp(\mathbf{w} \cdot \Phi(y, \langle b_i, b_j \rangle))}$$

絶対モデルの場合, 素性抽出関数 $\Phi(\cdot)$ は, y にも依存する形となる[2]. さらに, 絶対モデルは係るか係らないかの二値で周辺化するのに対し, 相対モデルは係り先候補集合 C_i で周辺化する点が異なる.

3.2 パラメータ推定

パラメータ (射影ベクトル) $\mathbf{w}$ は一般的な最尤推定を用いてを選択することができる. つまり, 学習データ $T = \{\langle b_i, d_i \rangle\}_{i=1}^{L}$ に対する対数尤度 $\mathcal{L}_\mathbf{w}$ の最大化を行う.

$$\begin{aligned}\hat{\mathbf{w}} &= \underset{\mathbf{w}}{argmax}\, \mathcal{L}_\mathbf{w} \\ \mathcal{L}_\mathbf{w} &= \sum_i \log(p(b_i \to b_{d_i} | C_i)) \\ &= \sum_i \log\Big(\sum_{c \in C_i} \exp\big[\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle) - \mathbf{w} \cdot \Phi(\langle b_i, c \rangle)\big]\Big)\end{aligned}$$

この時, 対数尤度を大きくするには, b_i の係り先候補 $c \in C_i$ について $\sum_{c \in C_i} \exp\big[\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle) - \mathbf{w} \cdot \Phi(\langle b_i, c \rangle)\big]$ を大きくすればよい. これは, まさしく**戦略 2** を近似的に実現することに他ならない.

最尤推定はしばしば過学習の問題を引き起こす. そこで, 過学習を防ぐためにパラメータの正則化を行う. これは**事後確率最大化 (MAP)** とも呼ばれ, パラメータの事前分布を考慮する最尤推定の一般形である. 事前分布を一様分布にすると, 通常の最尤推定と同一になる. 本稿では, Gaussian Prior (L2-norm) [1] の事前分布を考える. 正則化を行った場合, 目的関数は以下のようになる.

$$\mathcal{L}_\mathbf{w} = \sigma \sum_i \log(p(b_i \to b_{d_i} | C_i)) - \frac{1}{2}||\mathbf{w}||^2 \qquad (3)$$

$\sigma \in \mathbb{R}^+$ は, ハイパーパラメータであり, モデルの複雑さと学習データに対する適用度をコントロールする[3]. σ は, 交差検定等の一般的なモデル選択手法で選択する.

最適解 $\hat{\mathbf{w}}$ は, IIS や GIS といった反復スケーリング法 [1, 9] や, L-BFGS[8] といった準ニュートン法をを用いて求めることができる.

4 関連研究

絶対モデル以外にも, これまでいくつかの係り受け解析モデルが提案されている. ここでは, 内元らの**「後方文脈モデル」** [13], 金山らの**「3 つ組/4 つ組モデル」** [18], 及び我々が以前提案した**「チャンキングの段階適用法」** [16] と本手法との関連性について述べる.

4.1 後方文脈モデル (内元 他)

内元らは, 後方文脈を考慮する日本語係り受け解析モデル「後方文脈モデル」を提案している [13]. 絶対モデルを出発点にしているが, { 係る, 係らない } かの二値分類ではなく, { 係る, 手前の文節に係る, 越えて遠くに係る } の三値分類として学習を行う.

$$p(\text{係る} | \langle b_i, b_j \rangle) = \frac{\exp(\mathbf{w} \cdot \Phi(y = \text{係る}, \langle b_i, b_j \rangle))}{\sum_{y \in \{\text{係る, 手前, 越える}\}} \exp(\mathbf{w} \cdot \Phi(y, \langle b_i, b_j \rangle))}$$

[2] 多値分類を最大エントロピー法で行う時は, クラス数 × 素性数の素性を改めて素性とする事が多い. 二値分類の時は, $\Phi(y, \langle b, c \rangle) = y \cdot \Phi(\langle b, c \rangle)$ とすることも可能である.

[3] σ は SVM におけるソフトマージンパラメータと同様の働きをする.

係る確率 $p(b_i \rightarrow b_j|\langle b_i, b_j\rangle)$ は, 前方, 後方の文脈の確率を統合することで与えられる.

$$p(b_i \rightarrow b_j|\langle b_i, b_j\rangle)^2 = p(\text{係る}|\langle b_i, b_j\rangle)\prod_{k=i+1}^{j-1} p(\text{越える}|\langle b_i, b_k\rangle)\prod_{k=j+1}^{m} p(\text{手前}|\langle b_i, b_k\rangle)$$

後方文脈モデルは, 解析時にある種の文脈情報を捉えることができる. ただし, 候補集合の独立性を仮定しながら 「係る」「越える」「手前」の 3 つに分類しているために, 絶対モデルの本質的な欠点は解決されない. また, 学習と解析の戦略, 手法が異なる (学習は三値分類, 解析は確率値の統合) ため, 学習時には出現しなかった他の影響を解析時に受ける可能性がある. 一方, 提案手法は, 学習と解析は同一の戦略 (相対的な係りやすさの比較) に基づいており, 他の影響を受けにくい.

4.2 3 つ組/4 つ組モデル (金山 他)

金山らのモデルは, HPSG を用いてあらかじめ係り先の候補を 2 つ, ないし 3 つに限定することから始まる. 文節 b_i の係り候補が $b_{i,1}, b_{i,2}, b_{i,3}$ に限定された時, 文節 b_i が $b_{i,j}$ $j = 1, 2, 3$ に係る確率 $p(b_i \rightarrow b_{i,j})$ は以下で与えられる.

$$p(b_i \rightarrow b_{i,j}) = p(j|\langle b_i, b_{i,1}, b_{i,2}, b_{i,3}\rangle) = \frac{\exp(\mathbf{w}\cdot\Phi(j, \langle b_i, b_{i,1}, b_{i,2}, b_{i,3}\rangle))}{\sum_{j'=1,2,3}\exp(\mathbf{w}\cdot\Phi(j', \langle b_i, b_{i,1}, b_{i,2}, b_{i,3}\rangle))}$$

学習時には, 候補が 2 つのモデル (3 つ組モデル) と 3 つのモデル (4 つ組モデル) がそれぞれ作成される. これは, 係り先候補を限定させ, 二値分類ないし三値分類器をそれぞれ独立に構築していることに対応する. このような手法を一般の $3, \ldots, k$ つ組モデルに拡張するには, $2, \ldots, k-1$ 値分類器をそれぞれ個別に構築する必要があり, データスパースネスの問題が生じる. 3 つ組/4 つ組のみに限定することは, データスパースネスの問題を抑えつつ複数の候補を考慮できるバランスの取れた手法と考えられる.

しかし, 3 つ組/4 つ組モデルの問題点は, 事前に候補を 2 つないし 3 つに限定しなければならない点にある. 本手法は金山らの方法と同様に学習時に複数の候補を考慮できる一方で, それらの候補を事前に限定する必要はない.

4.3 チャンキングの段階適用 (工藤 他)

チャンキングの段階適用法 [16] は, 式 1,2 のように複数の候補から係り先を選択するような定式化になっていない. 代わりに, Shift-Reduce 法の一種に従い, 決定的に係り関係を同定する. 各ステップの動作手順 (現時点で shift するか reduce するか) を二値分類器を用い学習する. Reduce 動作は「係る」, Shift 動作は「係らない」とみなせるので一種の絶対モデルになっている.

チャンキングの段階適用法は, 近い文節に係りやすいという日本語係り受けの特徴をうまく活かした解析手法である. しかし, 後方の文脈を一切考慮しないため, 長距離の係り受けに弱くなる可能性がある.

5 実験および考察

実データを用い, 提案法と従来法の比較を行う. 比較対象は以下の 4 つである.

1) 相対モデル, 2) 絶対モデル [12, 14, 7], 3) 後方文脈モデル [13], 4) チャンキングの段階適用 [16].

金山らの 3 つ組/4 つ組モデル [18] は, 事前に候補を限定する必要があり, 評価用コーパスのみを用いての公平な比較が行えないために, 実験の対象外とした.

5.1 実験環境, 設定

京大コーパス (Version 3.0)[17] を以下の 3 つに分割して実験を行った.

- 学習データ: 一般記事 1 月 1,3-11 日 , 社説 1-8 月, 合計 24,263 文, 234,474 文節
- ディベロップメントデータ: 一般記事 1 月 12,13 日 , 社説 9 月, 合計 4,833 文, 47,580 文節
- テストデータ: 一般記事 1 月 14-17 日 , 社説 10-12 月, 合計 9,278 文, 89,982 文節

まず, 相対モデル, 絶対モデル, 後方文脈モデルに関する実験設定を説明する.

学習に用いた基本素性を表 1 に示す. これらは 若干な差異はあるものの [12, 14, 13, 7, 16] 等で用いられた素性であり, 日本語係り受け解析に用いられる素性として一般的なものである. ただし, **主辞**とは文節内で品詞が特殊, 助詞, 接尾辞となるものを除き, 文末に一番近い形態素, **語形**とは文節内で品詞が特殊となるものを除き, 文末に一番近い形態素のことを指す. さらに, 係り関係の情報を素性として与える**動的素性**も一部用いている. 動的素性の詳細は文献 [16] を参照されたい.

最大エントロピー法は線形分類器であるため, 素性の組みが重要な場合は, それらを明示的に与えなければならない[4]. 本稿では, 文献 [14] を参考に, 有効と思

[4] この説明は, 厳密には不正確である. Gaussian Prior を用いる場合は, Kernel 化が可能であり, 非線形モデルを原理的には構築可能である

表 1: 使用した基本素性

前/後文節	主辞見出し, 主辞品詞, 主辞品詞細分類, 主辞活用, 主辞活用形, 語形見出し, 語形品詞, 語形品詞細分類, 語形活用, 語形活用形, 括弧の有無, 句読点の有無, 文節の位置 (文頭, 文末)
文節間	距離 (1,2-5,6 以上), 括弧, 句読点の有無

われる素性の組みを人手で選択し, 新たな素性として投入した. また, 学習コーパス中に 3 回以上出現した素性のみを用いて実験を行った. 式 (3) におけるハイパーパラメータ σ は, ディベロップメントデータを用いて選択した.

解析手法として, 関根らの文末の文節から係り先を同定するアルゴリズム [15] を採用した. 関根らの手法では, ビームサーチを行いながら, 最良の解析木を導出する. 一方, ビーム幅を大きくしても必ずしも精度が向上するわけではなく, 場合によっては精度が低下すること, また, 決定的に解析しても同程度の精度が得られることが過去の研究で報告されている [13, 7]. そこで, ビーム幅は 1 とし, 決定的な解析を行った. すなわち, 文末の文節から, 式 (1),(2) を用いて最尤の係り先を決定的に選択していく.

チャンキングの段階適用法については, 既存システム CaobCha [5] をそのまま用いた. このシステムは, 1) SVM を学習, 分類手法として使ってること, 2) カーネル法を用い素性のカバレッジが高いこと, 3) 全ての動的素性を使っていること, の 3 点を考えると, 学習や素性の点で若干有利なシステムとなっている. SVM のソフトマージンパラメータ C は, ディベロップメントデータを用いて選択した.

なお, すべての実験は XEON 2.8Ghz, 主記憶 4Gbyte の Linux 上で行った.

5.2 実験結果

提案手法 (相対モデル) と, 従来方法 (絶対モデル, 後方文脈モデル, チャンキングモデル) の結果を表 2, 3 にまとめる. ただし, 係り受け正解率とは, 文末の一文節を除くすべての文節に対して, 正しく係り先が同定できたものの割合, 文正解率とは, 文全体の解析が正しいものの割合を示す.

同一データを用いてテストを行ったため, 出力は, 文節/文毎に対応が取れている. そこで, 対応が取れている場合の母比率の差を比較する手法であるマクネマー検定 [4] を用い, 個々のモデルの有意差を検証した. 検定では, 「母比率に差はない」という帰無仮説を立てる. P 値は, 帰無仮説が誤りでない確率値を意味する. 任意の有意水準 α (例えば 1%) を与え, $P \leq \alpha$ のとき, 帰無仮説を棄却し 「母比率に差がある」と結論付ける. 表 5 に マクネマー検定により得られた P 値を示す.

表 5: システム間の比較

システム	P 値 係り受け	P 値 文
相対 vs 絶対	1.3×10^{-12}	6.4×10^{-9}
相対 vs 後方	0.00014	0.0031
後方 vs 絶対	0.011	0.0012
相対 vs チャンキング	0.25	0.41
後方 vs チャンキング	0.10	0.21
絶対 vs チャンキング	0.00010	0.00050

結果, 相対モデルは絶対モデルと比較して有意水準 1% 未満で有意な差があることが分かる. 後方文脈モデルとの比較では, 文正解率の差は無いものの, 係り受けに関しては有意差が認められる. さらに, チャンキングモデルとは同等の性能だということが分かった. 学習アルゴリズムや素性の不利な点を考えると, 充分に高い性能だと考える. 絶対モデルと後方文脈モデルの有意差は認められない.

少なくとも, 相対モデルは絶対モデルに比べ有意に解析精度が高いことが言える. 後方文脈モデルに対しても, 他のデータセットを使った充分な検証が必要であるが, 精度が高いと言えるであろう. ただし, チャンキングモデルに対する優位性は現段階では結論付けられない.

5.3 距離毎の評価

相対モデルおよび後方文脈モデルは, 後方の文脈を含め全係り先候補を考慮するため, 絶対モデルやチャンキングモデルに比べ長距離係り受けの性能が高いと予想される. そこで, 係り先の距離毎に係り受け精度を算出し, 各モデルについて比較を行なった. 表 4 に距離毎の係り受け F 値を示す. ただし, 距離 n の係り受けの「精度」とはシステムが出力した距離 n の係り受けのうち, システムが正解した割合, 「再現率」とは正解データにある距離 n の係り受けのうち, システムが正解した割合である. F 値は, 精度と再現率の調和平均で定義される.

表 4 に示す結果は興味深い. 比較的距離が短かい (1-5) 場合は, チャンキングモデルの性能が他に比べ

[5] http://chasen.naist.jp/~ taku/software/cabocha/

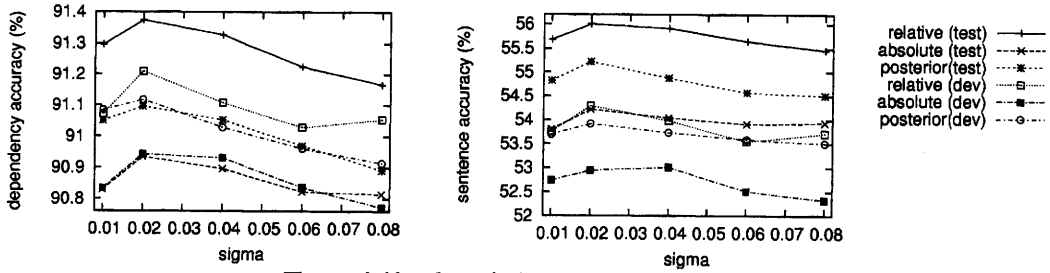

図 3: σ と係り受け正解率 (左), 文正解率 (右) の関係

表 2: 実験結果 (ディベロップメントデータ)

モデル	係り受け正解率 (%)	文正解率 (%)
相対モデル ($\sigma = 0.02$)	91.20 (38989/42747)	54.29 (2624/4833)
絶対モデル ($\sigma = 0.02$)	90.94 (38875/42747)	52.94 (2559/4833)
後方文脈モデル ($\sigma = 0.02$)	91.11 (38950/42747)	53.92 (2606/4833)
チャンキングモデル ($C = 0.001$)	**91.22** (38994/42747)	**54.36** (2627/4833)

表 3: 実験結果 (テストデータ)

モデル	係り受け正解率 (%)	文正解率 (%)
相対モデル ($\sigma = 0.02$)	**91.37** (73733/80695)	**56.00** (5201/9287)
絶対モデル ($\sigma = 0.02$)	90.93 (73379/80695)	54.21 (5035/9287)
後方文脈モデル ($\sigma = 0.02$)	91.09 (73510/80695)	55.21 (5128/9287)
チャンキングモデル ($C = 0.001$)	91.23 (73624/80695)	55.59 (5163/9287)

表 4: 係り先距離毎の比較: F 値, 精度/再現率 (テストデータ)

モデル	1	2-3	4-5	6-7	8-9	10 以上
相対モデル	97.2	86.7	78.1	**76.8**	**75.3**	80.8
	(96.8/97.6)	(88.7/84.6)	(76.7/79.6)	(77.4/76.2)	(75.3/75.3)	(79.1/82.5)
絶対モデル	97.1	85.5	77.0	75.1	74.6	80.7
	(96.3/97.9)	(89.6/81.8)	(75.0/79.2)	(76.8/73.4)	(74.3/74.9)	(76.9/84.8)
後方文脈モデル	97.0	85.9	78.0	76.2	74.9	**81.3**
	(96.4/97.7)	(88.7/83.2)	(76.8/79.3)	(76.6/75.8)	(74.5/75.3)	(79.3/83.5)
チャンキングモデル	**97.3**	**86.8**	**78.5**	75.3	72.6	79.4
	(97.1/97.5)	(88.5/85.2)	(78.6/78.4)	(73.9/76.8)	(71.3/74.0)	(76.0/83.2)

高い. 一方, 距離が 6 を越えると極端に F 値 (特に精度) が低下し, 相対モデルや後方文脈モデルといった全体の候補を考慮するモデルの性能が高くなっている. チャンキングモデルは, 近くの文節に係りやすいという性質を利用しているため, 距離の短かい係り受けが強くなっていると考察される. また, バックトラックを行わないため, 近距離の係り受けと判定されず長距離の候補としてまちがえて残ってしまう可能性がある. これらが係り受けの精度 (precision) を下げている要因であろう.

距離毎に性能差が観察される一方で, 係り受け距離が長い文節対の数は少ないために, 全体の平均を取ってしまうと, 双方のモデルに顕著な差がなくなってしまう. 近くに係りやすいという性質を重視するか, 全体を考慮するかは, トレードオフの関係にあることが改めて確認できた.

5.4 パラメータ σ の影響

図 3 に正則化パラメータ σ と正解率の関係を示す. テスト, ディベロップメントともに, どのモデルについても, $\sigma = 0.02$ の時に最良の正解率を示している. この結果, σ は係り受け解析に対して敏感なパラメータではなく, 交差検定等の一般的な方法で十分推定可能であることが分かった.

表 6: 素性数と学習時間

モデル	時間 (分)
相対モデル	71
絶対モデル	240
後方文脈モデル	402
チャンキングモデル	1009

また, 相対モデルは, σ の値によらず他の手法 (絶対モデル, 後方文脈モデル) に比べ高い正解率を示している.

5.5 学習時間の比較

絶対モデル, 後方文脈モデルは, 多値分類を元にしているために, 対立する係り関係の存在を考慮すると学習が困難になると予想される. 表 6 に, それぞれのモデルの学習時間を示す. 最大エントロピー法の学習は, 準ニュートン法の一種である L-BFGS[8] を用いて行った. SVM は SMO に基づく一般的な学習パッケージ[6]を用いている. 表 6 より, 相対モデルは学習効率という観点から見て他の手法より優れていると言える.

6 おわりに

本稿では, 「相対モデル」と呼ばれる日本語の統計的係り受け解析手法を提案した. 従来手法では, 着目している二文節のみから算出される絶対的な係りやさに基づき解析が行われていた. 一方, 日本語の係り受け解析は, 係り先候補から正解の係り先を 1 つだけ選ぶタスクである. そのため, 絶対的な係りやすさに基づき係り先を決定するよりは, 候補間での係りやすさの相対的な大小関係を比較するほうがタスクの性質をうまく反映している. 相対モデルは係りやすさの相対的な「差」に着目し, 学習を行う.

実データを用いた実験により, 相対モデルは解析精度の点で従来法に比べ同等かそれ以上の性能を示す一方で, 学習効率を大幅に改善できることが分かった.

参考文献

[1] Stanley F. Chen and Ronald. Rosenfeld. A gaussian prior for smoothing maximum entropy models. Technical report, Carnegie Mellon University, 1999.

[2] Michael Collins and Nigel Duffy. New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron. In *In Proc. of ACL*, pages 263–270, 2002.

[3] Yoav Freund, Raj D. Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.

[4] L. Gillick and S. Cox. Some statistical issues in the comparison of speech recognition algorithms. In *In Proc. of ICASSP*, pages 532–535, 1989.

[5] R. Herbrich, T Graepel, P Bollmann-Sdorra, and K Obermyer. Learning preference relations for information retrieval. In *ICML-98 Workshop: Text Categorization and Machine Learning*, 1998.

[6] Thorsten Joachims. Optimizing search engines using clickthrough data. In *In Proc. of SIGKDD*, 2002.

[7] Taku Kudo and Yuji Matsumoto. Japanese dependency structure analysis based on support vector machines. In *In Proc. of EMNLP/VLC*, pages 18–25.

[8] Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Programming*, 45(3, (Ser. B)):503–528, 1989.

[9] Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):380–393, 1997.

[10] Libin Shen and Aravind K. Joshi. An svm-based voting algorithm with application to parse reranking. In *In Proc. of CoNLL 2003*, pages 9–16, 2003.

[11] 磯崎秀樹, 賀沢秀人, and 平尾努. 優先度学習を用いた自然言語処理. In **情報処理学会研究報告** *2004-NL-161*, pages 105–110, 2004.

[12] 春野雅彦, 白井 諭, and 大山芳史. 決定木を用いた日本語係り受け解析. **情報処理学会論文誌**, 39(12):3117, 1998.

[13] 内元 清貴, 村田 真樹, 関根 聡, and 井佐原 均. 後方文脈を考慮した係り受けモデル. **自然言語処理**, 7(5):3–17, 2000.

[14] 内元 清貴, 関根 聡, and 井佐原 均. 最大エントロピー法に基づくモデルを用いた日本語係り受け解析. **情報処理学会論文誌**, 40(9):3397–3407, 1999.

[15] 関根 聡, 内元 清貴, and 井佐原 均. 文末から解析する統計的係り受け解析アルゴリズム. **自然言語処理**, 6(3):59–73, 1999.

[16] 工藤 拓 and 松本 裕治. チャンキングの段階適用による日本語係り受け解析. **情報処理学会論文誌**, 43(6):1834–1842, 2002.

[17] 黒橋 禎夫 and 長尾 眞. 京都大学テキストコーパス・プロジェクト. In **言語処理学会 第 3 回年次大会**, pages 115–118, 1997.

[18] 金山 博, 鳥澤 健太郎, 光石 豊, and 辻井 潤一. 3 つ以上の候補から係り先を選択する係り受けモデル. **自然言語処理**, 7(5):71–91, 2000.

[19] 飯田龍, 乾健太郎, and 松本裕治. 文脈的手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定. **情報処理学会論文誌**, 45(3):906–918, 2004.

[6] http://chasen.naist.jp/~ taku/software/tinysvm/