

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[†] 松本 裕治^{††}

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2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、2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大小関係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[†] and YUJI MATSUMOTO^{††}

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 classification 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* pairs of segments into two classes, but to *select* the most likely modifier 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、2文節間の係りやすさを決定するルールを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^{1)~6)}。

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2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といったような機械学習手法がこれまで適用されてきた^{1),2),5)}。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、2文節のみが与えられ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大小関係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。

[†] NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation

^{††} 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

相対モデルは、優先度学習⁷⁾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、複数の候補から正解を1つ選んだり、候補をランキングしたりするために提案された学習手法である。優先度学習は情報検索結果のランキング⁷⁾⁻⁹⁾、照応解析^{10),11)}、構文解析結果のランキング¹²⁾⁻¹⁴⁾、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とおりである。2章で日本語係り受け解析の概要に触れ、3章で従来法と提案手法の違いを述べる。4章で優先度学習に関するサーベイを行い、5章で具体的な学習アルゴリズムについて述べる。5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7章で京大コーパスを用いた評価実験を提示し、最後に8章で本稿をまとめる。

2. 統計的日本語係り受け解析

日本語の文に対し、その文節列を $B = \{b_1, b_2, \dots, b_m\}$ 、係り受けパターン列 $D = \{d_1, d_2, \dots, d_m\}$ 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。たとえば、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 に係るとき、 $b_i \rightarrow b_j$ 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1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 $S = \{\langle B_1, D_1 \rangle, \dots, \langle B_M, D_M \rangle\}$ を用い、入力文節列 $B \in \mathcal{B}$ から係り受けパターン列 $D \in \mathcal{D}$ 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。

日本語係り受けの手法として大きく2つがある。

A 確信度に基づく解析手法^{1),2),5)}

個々の文節の係り関係はすべて独立だと仮定し、統計的な確信度に基づき係り先をほかとは独立に1つ選択する。確信度の計算に、決定木、最大エントロピー法、SVM等が用いられている。解析時にはCYKといった一般的な構文解析手法や日本語の係り受け解析に特化した関根らの手法¹⁵⁾等が適用される。

B 決定的な解析手法⁶⁾

Shift-Reduce法の一つ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。各ステップは完全な二値であるために、Aのような確信度は必ずしも必要ではない。手法が単純であるにもかかわらず、確信度に基づく手法と同程度の解析精度が得られることが報告されている⁶⁾。

本稿では、Aの確信度に基づく手法を中心に議論を進める。Aの手法では個々の文節の係り関係は独立だと仮定するため、文節 b_i とその係り先文節番号 d_i のペア $T = \{\langle b_1, d_1 \rangle, \dots, \langle b_L, d_L \rangle\}$ 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 $C_i = \{b_{i+1}, \dots, b_m\}$ と表記する。さらに、2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、2文節の周辺のコンテキスト、あるいはそれらの組合せが素性ベクトルとして表現される。

3. 絶対モデルと相対モデル

2章で日本語係り受け解析には大きく2つの手法があることを示した。本稿では、確信度に基づく手法に焦点を当てる。この手法の論点はいかにして精度の高い確信度が算出できるかにある。まず、従来用いられていた絶対モデルとその問題点に触れ、次に本稿で提案する相対モデルについて述べる。

3.1 絶対モデル(従来法)

絶対モデルでは、候補となる2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 $f: \mathbb{R}^n \mapsto \{+1, -1\}$ 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等が絶対モデルに適用されている^{1),2),5)}。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 $y = \mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ 、 $\mathbf{w} \in \mathbb{R}^n$ 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。

$$\forall i, \forall c \in C_i$$

$$\mathbf{w} \cdot \Phi(\langle b_i, c \rangle) > 0 \text{ if } c = b_{d_i}$$

$$\mathbf{w} \cdot \Phi(\langle b_i, c \rangle) < 0 \text{ if } c \neq b_{d_i} \quad \square$$

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 $\text{sgn}(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))$ で与えられる。また、係りやすさの確信度は $\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ で近似できる。文節

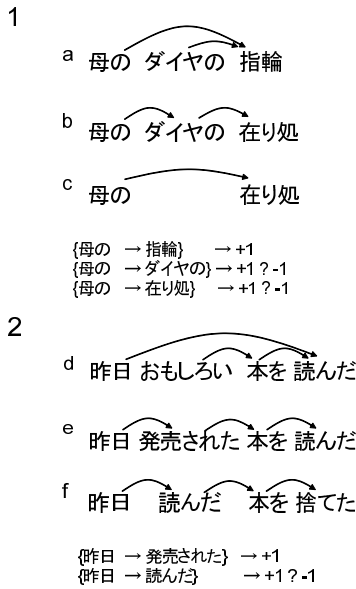


図 1 絶対モデルで学習困難な例

Fig. 1 Examples which cannot be handled with absolute model.

b_i の係り先 \hat{c}_i は以下で与えられる .

$$\hat{c}_i = \arg \max_{c \in C_i} \mathbf{w} \cdot \Phi(\langle b_i, c \rangle) \quad (1)$$

3.2 絶対モデルの問題点

絶対モデルは、候補 2 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妥当な手法なのであろうか .

図 1 に、絶対モデルでは学習が困難な 2 つの事例を示す . 文 a, b, c の係り関係 {母の \rightarrow ダイヤ}, {母の \rightarrow 在り処} は、正例、負例となる場合があり、対立する事例となってしまふ . d, e, f の係り関係 {昨日 \rightarrow 読んだ} も同様に対立する . このように、2 文節 $\langle b_i, b_j \rangle$ のみが与えられ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ある . 図 1 のような対立事例は、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 . このような事例を正しく解析するには、より広範囲の文脈を素性として投入する必要がある . ただし、やみくもに素性を投入すると精度が下がる場合があるため、素性の選択は慎重に行う必要がある . また、単純な手法として、対立事例に対してロバストな二値分類器を用いたり、なんらかの方法で対立事例を消去したりすれば上記の問題を解決できるかもしれない . しかし、裏を返せば絶対モデルがうまく働かば、二

値分類器そのものの性能や対立事例の消去法に依存することになるであろう .

さらに、絶対モデルは解析時において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 $\text{sgn}(\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle))$ 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 1 つも存在しなかったりした場合、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確信度を表現しているとは必ずしもいえない .

3.3 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 $C_i = \{b_{i+1}, \dots, b_m\}$ 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、係りやすさはほかとは独立に絶対的な基準で決まると考えるのではなく、他の候補との相対的な関係で決まると考えるほうがタスクの性質をうまく反映している . これ相対モデルの基本的なアイデアである .

図 1 文 a, b, c の例では、個々の 2 文節がそれぞれ係るかどうかを学習するのではなく、{母の \rightarrow 指輪} > {母の \rightarrow ダイヤ} > {母の \rightarrow 在り処} といった名詞間の係りやすさの嗜好性こそが本質的に学習すべき性質である . 文 d, e, f の例では、「昨日」はできるだけ近い過去形の動詞に係るという嗜好性を学習する必要がある . 「できるだけ近い」という嗜好性は他の候補と比較することで初めて成立するものであり、絶対モデルで扱うことは難しい .

上記のように、相対モデルは係りやすさの相対的な嗜好性を学習する . 絶対モデルで対立する事例集合は、相対モデルでは、対立していると考えのではなく、他候補と比べて係りやすさが大きかった (もしくは小さかった) と解釈され対立事例とならない . 図 1 の文集合から、上記のような嗜好性の大小関係を導出することは、二値分類器を導出する手続きに比べれば対立事例がないぶん容易である . もちろん、相対モデルを使っても対立する (大小関係が逆転する) 事例は存在する . しかし、その数は絶対モデルの対立数に比べれば非常に小さいであろう . また、相対モデルで対立する場合は、絶対モデルにおいても対立するため、相対モデルの対立事例数は絶対モデルのそれに比べて必ず

「昨日」といった副詞句の係り先は本質的に曖昧であるが、議論を簡単にするため 3 文の例から導出されるルールとしてこのような嗜好性を与えた .

少なくなる。

具体的には、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。

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

学習戦略 1 と学習戦略 2 の違いに注意されたい。学習戦略 1 では、各 2 文節を正負の二値分類していた。学習戦略 2 では、正解の文節ペア $\langle b_i, b_{d_i} \rangle$ の射影後の値（係りやすさ） $\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle)$ 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、3.2 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。

$$\hat{c}_i = \arg \max_{c \in C_i} \mathbf{w} \cdot \Phi(\langle b_i, c \rangle) \quad (2)$$

文節 $\langle b_i, b_j \rangle$ の係りやすさの確信度は $\mathbf{w} \cdot \Phi(\langle b_i, b_j \rangle)$ で与えられる。

4. 相対モデルと優先度学習

言語処理に限らず多くの分野で多値分類で定式化できるタスクとは別に、複数の候補から正解を 1 つ選択したり、候補をランキングしたりするタスクが存在する。後者のタスクは優先度学習で定式化するほうが好ましい。優先度学習は、Herbrich らによって定式化され⁷⁾、一般に学習戦略 2 の形をしている。相対モデルは優先度学習の日本語係り受け解析への自然な適用になっている。

二値分類手法として有名な Support Vector Machines (SVM)¹⁶⁾ や Boosting¹⁷⁾ に対応する優先度学習手法として、これまで Support Vector Ordinal Regression (SVOR)⁸⁾、Ranking SVM⁹⁾、Rank-Boost¹⁸⁾ 等が提案されている。また、最大エントロピー法¹⁹⁾ やその特殊形である Conditional Random Fields (CRF)²⁰⁾ も、広い意味で優先度学習と解釈することができる。

優先度学習と二値分類の本質的な違いは基準点の有無である。二値分類は正例、負例を分離する基準点（通常は 0）を設定し、基準点に対する大小関係に基づき事例を 2 つのグループに分類する。一方、優先度学習にはそのような基準点は存在せず、他と比べたときの相対的な大小関係が学習の対象になる。

優先度学習は多くの分野に適用されている。情報

検索の分野では、解析結果のランキングに SVOR や Ranking SVM が適用されている⁷⁾⁻⁹⁾。また、質問応答の分野では候補のランキングに最大エントロピー法が適用され、二値分類より精度が良いとの報告がある²¹⁾。Collins は、RankBoost、最大エントロピー法を構文解析結果のランキングに用いている¹²⁾。Ranking SVM を用いた構文解析結果のランキング手法もすでに提案されている^{13),14)}。磯崎らは、文の主辞の同定に Ranking SVM を適用し、二値分類より高性能であったことを示している¹¹⁾。飯田らは、トーナメントモデルと呼ばれる優先度学習手法を提案し、照応解析に適用している¹⁰⁾。トーナメントモデルの基本的な動機付けは優先度学習のそれと同一である。ただし、トーナメントモデルは 2 つの候補のどちらが良いかを通常の二値分類器を用いて学習するため、厳密には学習戦略 2 のような定式化にはなっていない。

5. 最大エントロピー法による定式化

4 章で、優先度学習の学習手法として大きく、Rank-Boost¹⁸⁾、SVM^{8),9)}、最大エントロピー法¹⁹⁾ があることを示した。本稿では、1) 従来の係り受け解析手法との比較に重点を置いていること、2) 学習の効率性、3) 規模耐性、の 3 点を考え、最大エントロピー法を採用する。

5.1 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 が与えられたとき、 b_i が $b_j (b_j \in C_i)$ 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 2) に用いられた手法と同一である。絶対モデルでは、文節ペア $\langle b_i, b_j \rangle$ が与えられたとき、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。

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

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。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

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

先候補集合 C_i で周辺化する点が異なる．

5.2 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．

$$\begin{aligned}\hat{\mathbf{w}} &= \arg \max_{\mathbf{w}} \mathcal{L}_{\mathbf{w}} \\ \mathcal{L}_{\mathbf{w}} &= \sum_i \log(p(b_i \rightarrow b_{d_i} | C_i)) \\ &= \sum_i \log \left(\sum_{c \in C_i} \exp[\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle) - \mathbf{w} \cdot \Phi(\langle b_i, c \rangle)] \right)\end{aligned}$$

このとき、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 $\sum_{c \in C_i} \exp[\mathbf{w} \cdot \Phi(\langle b_i, b_{d_i} \rangle) - \mathbf{w} \cdot \Phi(\langle b_i, c \rangle)]$ を大きくすればよい．これはまさに学習戦略 2 を近似的に実現することにほかならない．

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、通常最尤推定と同一になる．本稿では Gaussian (L2-norm)²²⁾ の事前分布を考える．正則化を行った場合、目的関数は以下ようになる．

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

$\sigma \in \mathbb{R}^+$ 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． 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^{22),23)} や L-BFGS²⁴⁾ といった準ニュートン法を用いて求めることができる．

6. 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」³⁾、金山らの「3つ組/4つ組モデル」⁴⁾、工藤らの「チャンキングの段階適用法」⁶⁾ と本手法との関連性について述べる．後方文脈、3つ組/4つ組モデルは確信度に基づく手法であり、チャンキングの段階適用法は決定的な解析手法である．

6.1 後方文脈モデル (内元ら)

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³⁾．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．

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

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．

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

後方文脈モデルは、解析時にある種の文脈情報をとらえることができる．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．

6.2 3つ組/4つ組モデル (金山ら)

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 $b_{i,1}, b_{i,2}, b_{i,3}$ に限定されたとき、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．

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

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 $3, \dots, k$ 組モデルに拡張するには、 $2, \dots, k-1$ 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

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．

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と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。

6.3 チャンキングの段階適用（工藤ら）

チャンキングの段階適用法⁶⁾は、前者2つの確信度に基づく手法と異なり、決定的な解析手法である。この手法では、Shift-Reduce法⁶⁾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」と見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、後方の文脈をいっさい考慮しないため、長距離の係り受けに弱くなる可能性がある。

7. 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、1) 相対モデル、2) 絶対モデル²⁾、3) 後方文脈モデル³⁾、4) チャンキングの段階適用⁶⁾、の4つである。

金山らの3つ組/4つ組モデル⁴⁾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実験の対象外とした。

7.1 実験環境、設定

京大コーパス（Version 3.0²⁾⁵⁾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日、3-11日、社説1-8月、合計24,263文、234,474文節
- デ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。これらは若干の差異はあるものの文献1)~3)、5)、6)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

表1 使用した基本素性
Table 1 Feature set.

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

す。さらに、係り関係の情報を動的に素性として与える動的素性⁶⁾も一部用いている。

最大エントロピー法は線形分類器であるため、素性の組が重要な場合は、それらを明示的に与えなければならない。本稿では、内元らが用いた素性²⁾を参考に、有効と思われる素性の組を手で選択し、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^{3)、5)}。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)もしくは(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CaboChaの学習モジュールをそのまま用いた。CaboChaは学習アルゴリズムとしてSVMを用いている。SVMのソフトマージンパラメータ C は、デベロップメントデータを用いて選択した。

素性設定や学習アルゴリズムが先の3つの手法（相対、絶対、後方モデル）とチャンキングの段階適用法で若干異なることに注意されたい。主な違いとして以下がある。

- 学習アルゴリズム：前者の3つは最大エントロピー法、後者はSVMを用いている。
- 組合せ素性：基本素性はそれぞれ同一である。た

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 2 実験結果

Table 2 Results of dependency accuracy and sentence accuracy.

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 3 システム間の比較

Table 3 Results of significance test.

システム 1 vs 2	P 値 (システム 1 のみが正解の数/システム 2 のみが正解の数)	
	係り受け	文
相対 vs 絶対	1.3×10^{-12} (1414/1060)	6.4×10^{-9} (487/321)
相対 vs 後方	0.00014 (1808/1585)	0.031 (594/521)
チャンキング vs 絶対	0.00048 (2565/2320)	0.0011 (820/692)
後方 vs 絶対	0.011 (1363/1232)	0.0012 (448/355)
チャンキング vs 後方	0.10 (2425/2311)	0.38 (764/729)
相対 vs チャンキング	0.11 (2415/2306)	0.34 (758/720)

だし、素性の組合せに関して、前者の 3 つは手動で展開している。後者は、多項式カーネルを用いることで自動的に組合せが展開される。

- 動的素性：動的素性の種類として、A) 係り元にすでに係る文節、B) 係り先にすでに係る文節、C) 係り先が係る文節、の 3 種類が存在する⁶⁾。チャンキングの段階適用法は、ボトムアップの Shift-Reduce 法を基にしており、スコープの狭い (係り距離の短い) ものから順番に係り関係が同定される。そのため、A, B, C すべての動的素性を用いることができる。相対、絶対、後方文脈モデルは、文末から各文節の係り先を 1 つずつ決めていく関根らの手法¹⁵⁾を用いているため、動的素性 A が使えない。

これらの相違のほとんどは学習手法のそれに起因する。SVM は最大エントロピー法に比べて高精度とされている点、多項式カーネルによる組合せ素性の自動展開が手動展開に比べカバレッジが高いという 2 点で、チャンキングの段階適用法が若干有利になっていると考えられる。

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

7.2 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 にまとめる。ただし、係り受け正解率とは文末の 1 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/文ごとに対応がとれている。そこで、対応がとれ

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²⁶⁾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 値は帰無仮説が真であるという仮定のもとに棄却してしまう確率である。任意の有意水準 α (たとえば 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 3 にマクネマー検定により得られた P 値を小さい順に示す。また、P 値とともに各システムのみが正解となった数も示している。たとえば、1 行目左の 1,414/1,060 は、相対モデルのみが正解となった数が 1,414、絶対モデルのみが正解となった数が 1,060 ということを表す。マクネマー検定はこの 2 つの数のみから P 値を算出する。

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、文正解率の差はないものの、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相対モデル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、十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係り受け正解率での有意差は認められない。

チャンキングモデルと他の 3 手法を詳しく比較すると、絶対的な相違数が 3 手法間の相違数に比べ約 1.7 倍程度大きい (1,500 前後 vs 2,500 前後) ことが分かる。つまり、相対モデルとチャンキングモデルは正解率における有意差は小さいが、出力結果そのものには大きな違いがあるといえる。

7.3 距離ごとの評価

相対モデルおよび後方文脈モデルは、後方の文脈を

表 4 係り先距離ごとの比較：F 値，(精度/再現率)

Table 4 Relation between dependency distance and accuracy (F-measure/precision/recall).

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

含め全係り先候補を考慮するため、長距離係り受けの性能が高く、チャンキングモデルは、直後に係りやすいという性質を利用しているため、短距離係り受けの性能が高いのではないかと予想される。このような解析手法の性質の相違が、表 3 における絶対的な相違数に現れていると考察される。そこで、係り先の距離ごとに係り受け精度を算出し、各モデルについて比較を行った。

表 4 に距離ごとの係り受け 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 において、各列の上段に F 値、下段に精度と再現率を提示している。

表 4 に示す結果は興味深い。比較的距離が短い(1-5)場合は、チャンキングモデルの性能がほかに比べ高い。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。この結果は、我々の予想と合致する。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。

7.4 品詞ごとの評価

表 5, 6 に、2 システム間の性能を品詞ごとに評価した結果を示す。表 5 は、相対モデルと絶対モデルの比較、表 6 は、相対モデルとチャンキングモデルの比較である。ただし、 P 値とはマクネマー検定に算出された値であり、有意差が大きいほど P 値が小さくなる。それぞれの表では、 P 値が 10%未満のもののみを列挙している。さらに、品詞とは、係り元の機能語(機能語が未定義の場合は主辞)の品詞大分類である。ただし、助詞のみ語彙化し、活用が定義されるものは活用も品詞に含めている。

表 5, 6 の結果から、相対モデルは動詞の連用形(連用節)の係り関係について有効に機能していることが

分かった。動詞の連用形の多くは連用中止表現であり、一般に係り先の判定は難しいとされている。その点から、この有意差は価値あるものだと考える。以下に連用節の係り関係の具体例を示す。下線が係り元、枠で囲った文節 1, 2 がそれぞれ相対モデル、絶対モデルが選んだ係り先である。絶対モデルは下記の例のように文末の文節を選択していることが多かった。

新民連問題の深刻化に 伴い、自党内の 反感が
¹ 強まっている ことを 浮き彫りに ² した。

その ためには、反核姿勢を 従来にも まして、
¹ 明確にする 必要が ² あろう。

絶対モデルは、位置素性(文末/文頭)を過大視することで、連用節の係り受け関係を半ば強制的に弁別していたものと考えられる。実際に正解データを調べたところ、連用節の係り受けの実に 46%が文末に係っていた。つまり、位置素性は弁別する目的だけを考えるときわめて有効に機能するといえる。

さらに、表 5 から係り先の同定の難しい品詞(助詞も、で、は、名詞、副詞)に対して相対モデルが有効に働いていることが分かる。助詞の「も」に関しても、連用節の考察と同様、以下の例のように文末に引っ張られる事例が多かった。

何事も ¹ なかったように 沖縄返還は ² 完了した。

一方、表 6 のチャンキングモデルとの比較を見ると、比較的簡単な品詞(助詞を、に、接続詞)についてモデルに差が出ていることが分かる。また、有意差はそれほど顕著ではないものの、名詞、副詞といった係り受けは、チャンキングモデルの方が高性能であることが分かった。名詞や副詞は、連用形の係り受けに比べ近距離性のバイアスを受けやすい(近距離の文節に係りやすい)ため、チャンキングモデルの精度が高くなっているものと考察される。

7.5 相対モデルとチャンキングモデルの組合せ

7.3 節で、相対モデルとチャンキングモデルは巨視

表 5 品詞ごとの比較 (1: 相対モデル, 2: 絶対モデル)

Table 5 Relation between part-of-speech and accuracy (1: relative 2: absolute).

品詞	正解数			精度		P 値	
	両方正解	1 のみ正解	2 のみ正解	両方正解	1 の精度		2 の精度
動詞-連用	2922	163	92	450	85.06	83.10	0.00001
助詞-も	1766	71	36	178	89.57	87.86	0.00101
名詞	3641	116	77	655	83.69	82.82	0.00623
助詞-に	6194	97	62	251	95.26	94.73	0.00701
助詞-は	5958	191	143	963	84.76	84.09	0.01012
助詞-で	2205	71	44	274	87.74	86.70	0.01533
助詞-が	5907	108	81	446	91.94	91.53	0.05856
副詞	2018	80	57	285	85.98	85.04	0.06016

表 6 品詞ごとの比較 (1: 相対モデル, 2: チャンキングモデル)

Table 6 Relation between part-of-speech and accuracy (1: relative 2: cascaded chunking).

品詞	正解数				精度		P 値
	両方正解	1 のみ正解	2 のみ正解	両方正解	1 の精度	2 の精度	
動詞-連用	2823	262	190	352	85.06	83.07	0.00084
助詞-を	7565	91	62	106	97.85	97.48	0.02356
助詞-に	6200	91	120	193	95.26	95.70	0.05391
接続詞	770	26	13	60	91.60	90.10	0.05466
助詞-は	5768	381	329	777	84.76	84.04	0.05562
副詞	1999	99	127	215	85.98	87.13	0.07249
名詞	3532	225	263	469	83.69	84.54	0.09395
助動詞-連体	230	3	10	10	92.09	94.86	0.09609

表 7 組合せの実験結果

Table 7 Results of model combination.

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)
チャンキングモデル ($C = 0.001$)	91.23 (73624/80695)	55.59 (5163/9287)
組合せ ($d = 3$, 交差戦略= b)	91.66 (73969/80695)	56.30 (5229/9287)

的な評価をすると差がないが、距離ごとに細かく評価すると性質が大きく変わることが分かった。この性質はデベロップメントデータの解析結果においても観察された。この結果から、2つのモデルを組み合わせ、互いの欠点を補うことでより高い正解率が得られるのではないかと考察される。

近距離の係り受けの性能は、長距離に比べ高い。そこで、チャンキングモデルが出力する係り先の距離が d 以下の場合は無条件にそれを採用し、それ以外は相対モデルの係り先を採用するという単純な組合せ手法を試みた。ただし、組合せにより非交差条件が崩れる場合は、a) チャンキングモデルを優先する、b) 相対モデルを優先する、の2つの場合を試みる。距離の閾値 d 、非交差条件の戦略 (a), (b) はデベロップメントデータを用いて選択する。

表 7 に組合せ結果を示す。7.2 節と同様の手法で有意差を検定したところ、文正解率では有意差は認められなかったものの、係り受け正解率では双方のモデルを統合した結果が個別の結果に対し有意な差があることが分かった。係り受けの確信度等は用いず、距離の

表 8 学習時間

Table 8 Training efficiency.

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

みを用いモデルを切りかえる単純な手法であるにもかかわらず、優位性を確認できたことは興味深い結果であると考えられる。

7.6 学習時間の比較

絶対モデルは二値分類、後方文脈モデルは多値分類を基にしているために、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 8 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つである L-BFGS²⁴⁾を用いて行った。SVM は一般的な学習パッケージを用いている。表 8 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いえる。

8. 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2文節のみから算出される絶対的な係りやす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大小関係に着目し、学習を行う。

実データを用いた実験により、以下の5点が確認できた。

- (1) 相対モデルは、従来法（絶対モデル、後方文脈モデル）に比べ高い係り受け正解率（91.37%）を示した。
- (2) 相対モデルは、決定的な解析手法（チャンキングの段階適用法）と同程度の性能であった。ただし、係り先の距離ごとに比較すると、前者は長距離依存に、後者は短距離依存に強いことが観察された。
- (3) 相対モデルは、従来法に比べ、連用節の係り受け等、解析が困難な事例に対する改善が顕著であった。
- (4) (2)の性質を考慮し、互いの欠点を補完するよう2つの手法の解析結果を組み合わせることで、さらに高い係り受け正解率（91.66%）を達成できた。
- (5) 相対モデルは、従来法に比べ学習効率が良いことが分かった。

統計的係り受け解析における研究では、広範囲、長距離の文脈を考慮しないと高精度は見込めず、いかにしてそれらの情報をモデルに反映するかという議論がなされてきた。しかし、局所的な情報のみで動く単純なチャンキングモデルと、広範囲の文脈を見るモデル（絶対、相対、後方文脈モデル）は、解析精度という観点でほぼ同一のパフォーマンスであるという事実が本実験を通じ明らかになった。これは一見逆説的である。つまり、モデルの定式化やその能力の議論抜きに、やみくもに広範囲の情報を投入するだけでは精度向上が見込めないことをこの事実は物語っている。従来の絶対モデルはその定式化の悪さから、文脈情報を十分に活かしてきれていなかった。広範囲の文脈を考慮するには、文脈情報が悪影響とならないような頑健かつ高

性能な機械学習モデルを適用/提案することが不可欠である。そのような意味で相対モデルは一步前進したのではないかと考える。

一方、相対モデルとチャンキングモデルとを詳細に分析すると、前者は長距離係り受けに、後者は近距離の係り受けに強いという事実が明らかとなった。すなわち、(1) 広範囲の文脈を見て解析する、(2) 近距離の文節に係りやすいという性質を使って解析する、という相反する2つの戦略のバランスをうまくとらないと、高い精度が実現できないということをこの実験結果は示唆している。本稿で示した2つの単純な組合せでも、比較的良好な精度が得られたことは、この事実の1つの裏付けになっていると考える。今後は、いつ、どのような状況で長距離文脈を見ればよいのか、逆に局所的な情報だけで解析できるのかといった議論を含め、これら2つの戦略を効果的に統合できる手法を提案したい。

参考文献

- 1) 春野雅彦, 白井 諭, 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, Vol.39, No.12, p.3117 (1998).
- 2) 内元清貴, 関根 聡, 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, Vol.40, No.9, pp.3397-3407 (1999).
- 3) 内元清貴, 村田真樹, 関根 聡, 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, Vol.7, No.5, pp.3-17 (2000).
- 4) 金山 博, 鳥澤健太郎, 光石 豊, 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, Vol.7, No.5, pp.71-91 (2000).
- 5) Kudo, T. and Matsumoto, Y.: Japanese Dependency Structure Analysis Based on Support Vector Machines, *Proc. EMNLP/VLC*, pp.18-25, (2000).
- 6) 工藤 拓, 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, Vol.43, No.6, pp.1834-1842 (2002).
- 7) Herbrich, R., Graepel, T., Bollmann-Sdorra, P. and Obermyer, K.: Learning Preference Relations for Information Retrieval, *ICML-98 Workshop: Text Categorization and Machine Learning* (1998).
- 8) Herbrich, R., Graepel, T. and Obermayer, K.: *Advances in Large Margin Classifiers*, MIT Press, chapter Large Margin Rank Boundaries for Ordinal Regression, pp.115-132 (2000).
- 9) Joachims, T.: Optimizing search engines using clickthrough data, *Proc. SIGKDD* (2002).
- 10) 飯田 龍, 乾健太郎, 松本裕治: 文脈の手がが

- 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, Vol.45, No.3, pp.906–918 (2004).
- 11) 磯崎秀樹, 賀沢秀人, 平尾 努: 優先度学習を用いた自然言語処理, 情報処理学会研究報告 2004-NL-161, pp.105–110 (2004).
- 12) Collins, M.: Discriminative Reranking for Natural Language Parsing, *Proc. ICML*, pp.175–182 (2000).
- 13) Collins, M. and Duffy, N.: New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures and the Voted Perceptron, *Proc. ACL*, pp.263–270 (2002).
- 14) Shen, L. and Joshi, A.K.: An SVM-based voting algorithm with application to parse reranking, *Proc. CoNLL 2003*, pp.9–16 (2003).
- 15) 関根 聡, 内元清貴, 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, Vol.6, No.3, pp.59–73 (1999).
- 16) Boser, B., Guyon, I. and Vapnik, V.: A Training Algorithm for Optimal Margin Classifiers, *Proc. 5th COLT*, pp.144–152 (1992).
- 17) Freund, Y. and Schapire, R. E.: Experiments with a new Boosting algorithm, *Proc. ICML* (1996).
- 18) Freund, Y., Iyer, R.D., Schapire, R.E. and Singer, Y.: An Efficient Boosting Algorithm for Combining Preferences, *Journal of Machine Learning Research*, Vol.4, pp.933–969 (2003).
- 19) Chellappa, R. and Jain, A.: *Markov Random Fields: Theory and Applications*, Academic Press (1993).
- 20) Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. ICML*, pp.282–289 (2001).
- 21) Ravichandran, D., Hovy, E. and Och, F.J.: Statistical QA — Classifier vs. Re-ranker: What's the difference?, *Proc. ACL 2003 Workshop on Multilingual Summarization and Question Answering*, pp.69–75 (2003).
- 22) Chen, S.F. and Rosenfeld, R.: A Gaussian prior for smoothing maximum entropy models, Technical report, Carnegie Mellon University (1999).
- 23) Pietra, S.D., Pietra, V.D. and Lafferty, J.: Inducing Features of Random Fields, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.4, pp.380–393 (1997).
- 24) Liu, D.C. and Nocedal, J.: On the limited memory BFGS method for large scale optimization, *Math. Programming*, Vol.45, No.3, (Ser. B), pp.503–528 (1989).
- 25) 黒橋禎夫, 長尾 眞: 京都大学テキストコーパス・プロジェクト, 言語処理学会第3回年次大会, pp.115–118 (1997).
- 26) Gillick, L. and Cox, S.: Some Statistical Issues in the Comparison of Speech Recognition Algorithms, *Proc. ICASSP*, pp.532–535 (1989).

(平成 16 年 6 月 29 日受付)

(平成 17 年 2 月 1 日採録)



工藤 拓 (正会員)

1999 年京都大学工学部電気電子工学科卒業。2001 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。2004 年同博士後期課程修了。同年より NTT コミュニケーション科学基礎研究所, リサーチアソシエイト。現在に至る。工学博士。2001 年度本学会山下記念研究賞受賞。統計的自然言語処理, テキストマイニング, 機械学習に興味を持つ。



松本 裕治 (正会員)

1977 年京都大学工学部情報工学科卒業。1979 年同大学大学院工学研究科修士課程情報工学専攻修了。同年電子技術総合研究所入所。1984～1985 年英国インペリアルカレッジ客員研究員。1985～1987 年(財)新世代コンピュータ技術開発機構に出向。京都大学助教授を経て, 1993 年より奈良先端科学技術大学院大学教授, 現在に至る。工学博士。専門は自然言語処理。人工知能学会, 日本ソフトウェア科学会, 言語処理学会, 認知科学会, AAAI, ACL, ACM 各会員。