

# 複数のパーザを利用した統計的部分係り受け解析

乾 孝 司<sup>†</sup> 乾 健 太 郎<sup>††</sup>

我々はこれまでに、信頼のおける部分だけを出力し被覆率を犠牲とする代償として正解率を向上させる統計的部分解析手法の調査を進めてきた。本稿では、さらにこの考えに委員会方式という概念を統合した枠組みを提案し、その評価を行った。委員会方式とは、複数の解析器(委員)の出力解を組み合わせることで解析精度の向上をはかる手法である。ここでは、各委員から得られる解析結果に基づき委員会で多数決により統計的部分解析を行うために、従来の基本的な委員会方式に次の3つの拡張を施した:(1) 解析器(委員)が推定した係り受け確率を票の重みと見なして重み付きの票を投じる確率的投票,(2) 委員間での票の重みの信頼性を標準化する重み標準化,(3) 各係り文節に対して2位以下の係り先候補にも重み付きの票を投じる多重投票。既存の5つの統計的解析器を用いて、京大コーパスを対象データとする解析実験を行った。その結果、委員の組合せによって精度変化には多少の揺れがあるものの、総合的には提案した枠組みおよび3つの拡張が解析精度の向上に有効に作用する見通しを得た。

## Committee-based Decision Making in Probabilistic Partial Parsing

TAKASHI INUI<sup>†</sup> and KENTARO INUI<sup>††</sup>

In this paper, we explored two new direction for the next step beyond the state of the art of statistical parsing: probabilistic partial parsing and committee-based decision making. Probabilistic partial parsing makes only as an output partial parse tree that is probabilistically highly reliable. Committee-based decision making is to combine several outputs from different systems (parsers) to make a better decision. Aiming at this coupling, we present a general framework which have three extensions against original basic framework to committee-based decision making. (1) **probabilistic voting**: a committee accepts probabilistically parameterized votes as its input. (2) **weight standardization**: a committee provides a means for standardizing original votes to guarantee reliability of them. (3) **multiple voting**: a committee allows a committee member to vote not only to the best-scored candidate but also to all other potential candidates. From the result of our experiments on the Kyoto Japanese corpus, we show that our presented framework have some contributions.

### 1. はじめに

日本語文の係り受け解析技術は、機械翻訳などの応用システム、あるいはテキスト言い換え技術などのより高次の解析にとって必要不可欠な要素技術であると認識されている。近年では、大規模な構文情報つき言語コーパスが構築されたこととともなって、統計情報に基づく係り受け解析手法が目覚ましい成果をあげており注目されている。しかしながら、係り受け解析という作業が実際には意味や談話などの情報と互いに依

存しあっていることを考えると、統計的手法を単独で用いるアプローチには深刻な限界があると予想される。実際に報告されている精度も90%前後を境にして伸び悩みの傾向を示している。

このように現状の解析精度は応用システム側から見れば不十分といわざるをえない。しかしながら、正解となりうる複数の出力を認める冗長解析<sup>(6)</sup>や、信頼のおける部分のみを出力する部分解析<sup>(6),13),15)</sup>などの手法が提案され、これらによって解析器の不完全な部分を補うことにより、係り受け解析がこれまでに以上に利用しやすくなってきている現状がある。

また、これまで係り受け解析の精度を向上させるためには、解析器あるいは言語モデルを洗練・修正するアプローチが多くとられてきた。その一方で、委員会方式のように解析器や言語モデル自体には直接言及せず、複数個の既存の資源をうまく利用して精度の向

<sup>†</sup> 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

<sup>††</sup> 九州工業大学大学院情報工学研究科  
Graduate School of Computer Science and Systems En-  
gineering, Kyusyu Institute of Technology

上をねらう、アンサンプルを考慮したアプローチも検討されている。たとえば Henderson ら<sup>12)</sup>は、複数の既存の英文構文解析器の出力解を多数決させることにより、用いたどの解析器よりも精度が向上したと報告している。委員会方式の適用に関する研究は、構文・係り受け解析に限らず、品詞タグづけ<sup>1),4),23)</sup>、多義性解消<sup>19)</sup>、機械翻訳<sup>8)</sup>、音声認識<sup>7)</sup>などの分野ですでに検討されており、いずれにおいてもその有効性が報告されている。また、委員会方式と類似した方式であるブースティング方式 (boosting<sup>9),20)</sup>も近年では注目を集めている。ブースティングでは、複数の弱仮説に対して例外事例に学習の重みを集中させていくことを行い、そして最終的な強仮説を弱仮説の重みつき多数決によって得ることを可能にしている。Haruno ら<sup>11)</sup>は、初期のブースティングにおける問題点を解消した AdaBoost 方式<sup>10)</sup>を決定木学習に適用し、日本語文の係り受け解析実験から高い解析精度を得たという興味深い結果を報告している。

上記のような委員会方式を実際に適用するには、その委員となる複数の解析器が必要となる。特に、統計的部分解析を仮定すれば複数の統計的係り受け解析器が必要となるが、近年では統計的解析器 (確率言語モデル) を構築する研究が数多く報告されており、委員として利用できる十分な数の解析器が存在している。本稿での我々の興味は、このような現状において、すでに利用可能な複数の解析器を複合的に利用することを考えた場合に、個々の解析器の出力解を統合するための方法論を経験的に検討することである。具体的には、統計的部分解析手法と委員会方式という2つの概念を統合した枠組みについて考察する。各委員から得られる解析結果に基づき委員会で多数決により統計的部分解析を行うために、従来から提案されている委員会方式に3つの確率的な拡張を施し、それらの拡張が解析精度の向上にどの程度有効に働くかを調査した。

以下、2章で統計的部分係り受け解析の概要を説明し、3章において委員会方式、およびその確率的一般化について述べる。4章では係り受け解析実験の結果とその考察を述べ、5章で本稿をまとめる。

## 2. 統計的部分係り受け解析

部分係り受け解析では、入力文に対応したすべての係り受け関係を出力するのではなく、信頼のおける (係り受け解析器が「自信を持っている」) 係り受け関係のみを選択的に出力する。近年では、統計的なアプローチによる係り受け解析手法がさかんに研究されている背景を受け、係り受け関係の信頼性を見積もる尺

度に統計的尺度を適用した統計的部分解析が提案されている。以下では4章で用いた統計的尺度の例として、係り受け確率の推定法の一例を紹介する。

### 2.1 係り受け確率

入力文を  $s$ 、その文節列を  $b_1 b_2 \dots b_n$  とし、 $b_i$  が  $b_j$  に係ることを  $r(s, i, j)$  で表すとす。  $s$  が係り受け関係  $r(s, i, j)$  を持つ確率を

$$P(r(s, i, j)|s) = \frac{P(s, r(s, i, j))}{P(s)} \quad (1)$$

$$\forall i. \sum_j P(r(s, i, j)|s) = 1$$

と定式化し、この確率を係り受け確率と呼ぶことにする。

個々の係り受け確率は、言語モデルに従って n-best の係り受け構造候補を求めたあと、各係り受け関係を含む候補の確率を出現係り受けごとに足し合わせ、n-best の候補の確率和で正規化することによって推定できる。

ここで、各  $b_i$  について係り受け確率を最大にする係り受け関係を  $r(s, i, g^*(s, i))$  とする。ただし  $g^*(s, i)$  は次式によって求められる。

$$g^*(s, i) = \arg \max_j P(r(s, i, j)|s) \quad (2)$$

すると、閾値  $\sigma$  よりも高い係り受け確率  $P(r(s, i, g^*(s, i))|s)$  を持つ係り受け関係  $r(s, i, g^*(s, i))$  を選択的に決定することで、部分解析が実現できる。

### 2.2 P-A 曲線

係り受け確率が  $P(r(s, i, g^*(s, i))|s)$  となる係り受け関係において、係り先が正しく決定された割合を示す係り受け正解率を  $A(P(r(s, i, g^*(s, i))|s))$  とする。係り先が正しく決定されるとは、テストセットとして採用したコーパスと係り先が等しいことを意味する。言語モデルに対して十分な量の学習データ (コーパス) を確保することが可能であれば、言語モデルが推定する係り受け確率と係り受け正解率の間には次の関係が成立する。

$$A(P(r(s, i, g^*(s, i))|s)) = P(r(s, i, g^*(s, i))|s)$$

ここで、係り受け確率 (dependency probability) と係り受け正解率 (accuracy) の関係をプロットして得られる曲線を P-A 曲線と呼ぶ。図1に示す3本の曲線は実験 (4章) から実際に得た P-A 曲線である。図より各解析器 (確率言語モデル) が出力する係り受け確率は、いずれもある程度正確に正解率を予測していることが分かる。

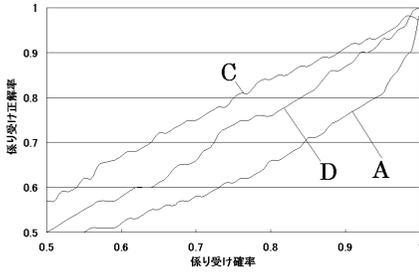


図1 P-A 曲線  
Fig. 1 P-A curve.

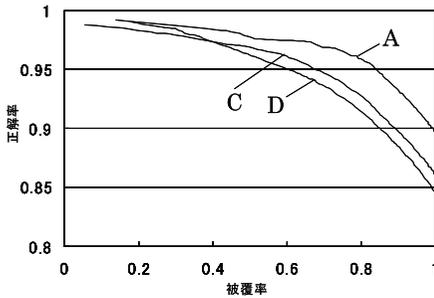


図2 C-A 曲線  
Fig. 2 C-A curve.

2.3 C-A 曲線

次に、 $r(s, i, g^*(s, i))$  の中で、閾値  $\sigma$  以上の確率を持つ係り受け関係だけを選択的に決定する作業を考える。  $\sigma$  を変化させた場合の被覆率 (coverage) と正解率 (accuracy) の関係をプロットして得られる曲線を C-A 曲線と呼ぶ。ただし、被覆率と正解率は次式によって求める。

$$\text{被覆率} = \frac{\text{係り先が決定された文節の数}}{\text{テストセット中の文節の数}} \quad (3)$$

$$\text{正解率} = \frac{\text{係り先が正しく決定された文節の数}}{\text{係り先が決定された文節の数}} \quad (4)$$

部分解析ではこの C-A 曲線が解析器 (確率言語モデル) の評価尺度となる。つまり、従来の係り受け解析では被覆率 = 1.0 に固定し、正解率のみでモデルの性能を評価していたが、部分解析では被覆率と正解率の組合せによってモデルを評価する。被覆率が同じなら正解率が高い方が望ましく、性能が良いことを示す。たとえば図 2 の例では A, C, D の順に性能が良いことを示している。以下、本稿では、被覆率 = 0.5, 0.55, ..., 1.0 の 11 点の平均係り受け正解率を 11-point 正解率と呼び、これを C-A 曲線の要約として用いる。また、11-point 正解率と区別するために、従来の被覆率 = 1.0 における正解率を総係り受け正解率と呼ぶことにする。

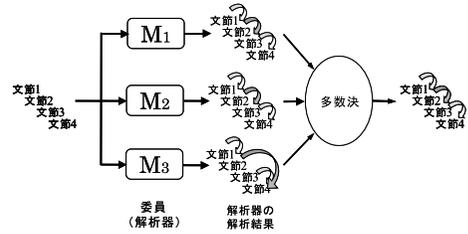


図3 単純な委員会方式の処理の流れ  
Fig. 3 Flow of simple framework of committee-based parsing.

3. 委員会方式の確率的一般化

3.1 委員会方式

委員会方式とは、複数の解析器 (委員) の出力解を組み合わせることで解析精度の向上をはかる手法である。係り受け解析にあてはめると、「入力文を複数の異なるパーザで解析させ、係り文節に対して各解析器が提出した係り先を多数決にかけ、そして最も票の集まった係り先を委員会としての意志決定結果として出力する」ことに対応する。

図 3 に処理の流れの例を示す。図中の例では、3 つの解析器が係り文節 2 の係り先としてそれぞれ文節 3、文節 3、文節 4 に投票しており、委員会は多数決 (文節 3 に 2 票、文節 4 に 1 票) の結果、文節 3 を文節 2 の係り先として出力している。

先述の例では重みなし多数決による単純な方法によって各解析器の係り先を混合していた。しかし、解析器として統計的係り受け解析器を仮定した場合には、解析器は任意の係り文節の係り先と同時にその係り受け関係の尤度を表す係り受け確率もあわせて出力することができる。本研究では、各委員から得られる係り受け確率に基づいて委員会で多数決により統計的部分解析を行うために、従来の委員会方式の拡張を試みる。

3.2 確率的一般化

図 4 に拡張を施した委員会方式の概要を示す。まず処理の流れを概観する。委員会は各解析器から出力される係り受け確率行列を入力行列として受け取り、重み標準化過程を経て、重み行列を得る。次に重み行列は、重みつき多数決過程に渡され、重みつき多数決過程において委員会としての意思決定を行う。そして最終的に 1 つの係り受け確率行列 (出力行列) を出力する。ここで、入力文  $s$  に対して係り受け確率  $P(r(s, i, j)|s)$  を  $i$  行  $j$  列要素を持つ行列のことを係り受け確率行列と呼んでいる。

従来の委員会方式に対する拡張点は次の 3 点である。(1) 各係り文節に対応する係り先に対して、係り受け

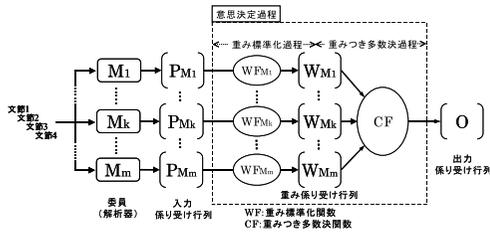


図4 確率的に一般化された委員会方式

Fig. 4 Committee-based probabilistic partial parsing.

確率を票の重みと見なして重みつきの票を投じる確率的投票。投票された重みつきの票は、重みつき多数決過程によって重みつき多数決がなされる。重みつき多数決過程の詳細は後述する。また、委員会での意志決定の結果より得られる解に対して部分解析を行うため、係り受け確率付きの出力ができるように拡張する。

(2) 重み標準化過程を導入し、委員間での票の重みの信頼性を標準化する重み標準化。既存の解析器の解析精度は90%前後であり、各解析器が内部的に推定する係り受け確率は必ずしも正確に係り受け正解率を予測するものではない。そこで、票の重みとしての係り受け確率を補正することで解析器間での重みの信頼性を保証する。重み標準化過程の詳細は後述する。

(3) 各係り文節に対して2位以下の係り先候補にも重みつきの票を投じる多重投票。解析器は、最も確率の高い1位の係り先候補に投票するだけでなく、2位以下のすべての係り先候補にも重みつきの票を投じる(入出力として係り受け確率行列を仮定する)。

3.2.1 重み標準化過程

図1のP-A曲線から分かるように、係り受け確率と実際の正解率の関係は用いる係り受け解析器によってばらつきがある。たとえば、図1の解析器Aが係り受け確率0.9と推定した場合と、解析器Cが同じく係り受け確率0.9と推定した場合では、実際の正解率に大きな開きがある。したがって、ある問題に対し両者がともに係り受け確率0.9と推定した場合にはCを信じるのが有利なことが分かる。このように、各解析器が推定する係り受け確率を混合する際には、何らかの補正を行い、解析器間での重みの信頼性を保証することが望ましい。この補正作業を重み標準化と呼ぶ。

次章で述べる実験では、以下に示す3つの標準化の戦略(重み標準化関数)を試験的に採用した。

Simple 入力行列をそのまま重み行列として用いる(後ろ2つの戦略との比較用)。

$$w_{ij}^{M_k} = P_{M_k}(r(s, i, j)|s) \tag{5}$$

ただし、 $M_k$  は委員会を構成する  $k$  番目の解析器、 $P_{M_k}(r(s, i, j)|s)$  は、入力文  $s$  を解析器  $M_k$  で解析した際に出力として得られる係り受け確率行列(入力行列)の  $i$  行  $j$  列要素の係り受け確率、 $w_{ij}^{M_k}$  は解析器  $M_k$  の出力に対応する重み行列  $W_{M_k}$  の  $i$  行  $j$  列要素を表す。

Normal 係り受け確率-係り受け正解率曲線(P-A曲線)が示す関数  $A_{M_k}$  を標準化に利用する。

$$w_{ij}^{M_k} = \alpha_i^{M_k} A_{M_k}(P_{M_k}(r(s, i, j)|s)) \tag{6}$$

ただし、 $A_{M_k}(p)$  は係り受け確率  $P_{M_k}(r(s, i, j)|s)$  を持つ係り受け関係の係り受け正解率を表す。また、 $\alpha_i^{M_k}$  は、補正された  $i$  行目の値を確率に正規化する(つまり  $\forall i, \sum_j w_{ij}^{M_k} = 1$  の関係を保証する)正規化関数である。

Class 委員会の構成委員となるそれぞれの解析器には問題クラスごとにその解決能力にばらつきがある。そのため、先述の標準化戦略 Normal で考慮する P-A 曲線では十分に正確な標準化ができない可能性がある。そこで問題クラスごとの P-A 曲線を標準化に利用する。

$$w_{ij}^{M_k} = \beta_i^{M_k C_{b_i}} A_{M_k C_{b_i}}(P_{M_k}(r(s, i, j)|s)) \tag{7}$$

ただし、 $A_{M_k C_{b_i}}(p)$  は訓練データ中に出現した問題クラス  $C_{b_i}$  に該当する係り受け関係における係り受け正解率を表す。今回は問題クラスの分類に係り文節  $b_i$  の係り属性などを利用している。また  $\beta_i^{M_k}$  は  $\alpha_i^{M_k}$  と同様の正規化関数である。

重みの標準化作業は、解析器(確率言語モデル)を洗練する作業に似ている。しかし、モデルの洗練では、誤った係り文節の係り先を正しい係り先へと変更することが期待されるのに対し、重み標準化作業では、解析を誤った係り文節の係り先に対して、誤りである係り先へはその重み(係り受け確率)をより小さく、正しい係り先へはその重みがより大きくなるように変更する。この作業は係り先を変更するほどの抜本的なものではないため、個々の解析器の解析精度の向上に直接的に貢献することは期待できない。しかし、誤った係り先への重みの増大を回避できるため、委員会での相補的な意志決定を行う際には効果的であると考えられる。

Normal, Class ともに、実際には過学習を避けるために獲得した P-A 曲線に適当なスムージングを施して用いている。

### 3.2.2 重みつき多数決過程

重み標準化過程より得られる重み行列の各要素は票の重みと見なされ、重みつき多数決過程で混合される。今回は重みつき多数決戦略（重みつき多数決関数）としては以下に示す Voting 戦略を採用した。 Voting 重みつき多数決を行う。

$$o_{ij} = \frac{1}{m} \sum_{M_k} w_{ij}^{M_k} \quad (8)$$

出力行列の要素は個々の重み行列の行列要素の平均値として計算される。ここで、 $o_{ij}$  は出力行列  $O$  の  $i$  行  $j$  列要素、 $m$  は委員会を構成する委員数を表す。

## 4. 評価実験

前章で述べた枠組みの有効性を検証するため、既存の代表的な係り受け解析器を構成委員とする委員会を作成して係り受け解析実験を行った。

### 4.1 セッティング

#### 評価実験に利用した解析器

委員会方式の委員となる解析器として下に示す5つの代表的な統計的係り受け解析器を用いた。各解析器はそれぞれ独立に作成されており、各々の特徴は大きく異なる。

- KANA<sup>5)</sup>: 最大エントロピー法に基づいたボトムアップモデル
- 茶掛<sup>6)</sup>: Collins モデル<sup>3)</sup>を拡張したボトムアップモデル
- SLUNG+QUADRUPLET<sup>16)</sup>: HPSG を利用したボトムアップモデル
- PGLR+LEX<sup>21)</sup>: 語の共起関係を加味したトップダウンモデル
- Peach Pie Parser<sup>22)</sup>: 最大エントロピー法に基づいたボトムアップモデル

PGLR+LEX, Peach Pie Parser に関しては、2.1 節で述べた方法により係り受け確率を推定し、重みと見なした。茶掛, SLUNG+QUADRUPLET に関しては、モデルが推定する部分分布  $P(r(s, i, j) | s)$  を直接係り受け確率として利用した。KANA は係り受け関係に確率ではないスコアを付与するため、スコアが  $[0.0, 1.0]$  の範囲をとるように適当なフィルタ関数を通して正規化した値を重みとして利用した。以下では、それぞれの解析器を匿名 (A ~ E のアルファベツ

表 1 評価実験に利用した解析器の単独での精度  
Table 1 The total/11-point accuracy achieved by each individual model.

解析器	総係り受け正解率	11-point 正解率
A	0.8973	0.9606
B	0.8733	0.9297
C	0.8613	0.9289
D	0.8466	0.9264
E	0.8089	0.8783
naive	0.5730	

ト)で参照する。後述する実験データによる個々の解析器の精度は表1のとおりである。

#### テストセット

実験データには、京大コーパス (ver.2.0)<sup>18)</sup>を用いた。まず、コーパスの全文 19,956 文を実験で用いる6つの解析器(先述の5つと後述の追実験で用いるKNP)で解析し、いずれかが解析に失敗した2,940文を取り除いた。

次に、それぞれの解析器が持つ文節区切りの認定規準が異なるために、残りの17,016文を対象として、細分されている文節をより大きな文節に合わせることで解析器間の文節区切りを統一化した。たとえば、ある単語列 ( $xyz$ ) に対して解析器 A が1文節 ( $b_{xyz}^A$ ) と認定し、解析器 B が3文節 ( $b_x^B, b_y^B, b_z^B$ ) と認定した場合は、3文節を統合して1文節と見なすことにした ( $b_x^B, b_y^B, b_z^B \rightarrow b_{xyz}^B$ )。さらに、統合する前に末尾以外に位置していたどの文節 ( $b_x^B, b_y^B$ ) も末尾文節 ( $b_z^B$ ) より後方には係らないという条件を課し、この条件にあてはまらない文節列を含む文は実験データから取り除いた。また、統合した後の文節 ( $b_{xyz}^B$ ) の係り先および係り受け確率は、統合する前に末尾に位置していた文節 ( $b_z^B$ ) の情報をそのまま利用した。以上の処理によって得られた14,234文を実験データとして用いた。

ここで、文節区切り統一化処理の結果、文節数が減少し (7.9 文節/文)、文節の頻度分布が変化していることに注意されたい。表2に頻度分布の変化の一例を示す。文節の分類は、4.2.3 項で用いる分類と同じものであり、たとえば、[動詞連体] は体言に連体修飾する動詞を含む文節を、[時相名詞 0] は時相名詞が含まれていて、読点を含んでいない文節をそれぞれ表している。また、図5に文節区切り変化の一例を示す。一

Voting 戦略のほかに、重み標準化の後に最も高い重みを持つ委員に従う Switching 戦略  $o_{ij} = \arg \max_{M_k} w_{ij}^{M_k}$  についても試行実験を行ったが、Voting よりも相対的に精度向上への貢献度が低い結果であった。

以下、評価実験を通して、文末の文節および文末から2番目の文節は評価対象外とした。

表2 文節区切り統一化による文節の頻度分布変化の一例  
Table 2 Examples of *Bunsetsu*-phrase frequency distribution.

問題クラス	統一化前	統一化後	減少率
[動詞連体]	5402	3753	30.5%
[の 0]	14103	13077	7.3%
[指示詞 0]	1751	1349	23.0%
[形容詞連体 0]	2550	2261	11.3%
[時相名詞 0]	713	89	87.5%
[副詞的名詞 0]	87	63	27.6%



図5 文節区切り統一化による文節区切り変化の一例  
Fig. 5 Examples of *Bunsetsu*-phrases boundary.

番上の例では、2つの文節「昨年」と「十月には」が1つの文節「昨年十月には」にまとめられたことにより、[時相名詞 0]に分類される文節が実験データから削除されている。近接する文節に係る事例が文節区切り統一化処理により削除されるため、実験データとして用いたデータは元のコーパスよりも係り受け推定問題の難しさが若干上がっている。各文節がすべて隣に係るとした場合の正解率は57.3%であった(表1下のnaive)。

上限値とベースライン

委員会方式では、係り文節に対して、複数の解析器から得られる係り先の中で最も多くの票を得た係り先を選択する作業を行う。このことは、正しい係り受けが委員のどの解析器からも支持されない(投票されない)場合には、正しい係り受けが委員会による意志決定の結果からは得られないことを意味している。

仮に2つの解析器( $M_1$ と $M_2$ )を委員とした場合(図6)、テストセット中の係り文節集合 $S$ 、それぞれの解析器が正しい係り先を推定できる係り文節集合 $M_1$ 、 $M_2$ に対し、塗りつぶされた部分集合 $M_1 \cup M_2$ が委員会による意志決定の結果から正しい係り受けが得られる可能性がある係り文節集合となる。本稿では、集合 $M_1 \cup M_2$ に含まれる係り文節の係り先を正しく推定することをタスクの上限として定め、テストセット中の係り文節集合全体に対するその割合を上限値とする。また、構成委員の中で最も解析精度の高い解析器が正しい係り先を推定できる割合(すなわち総係り

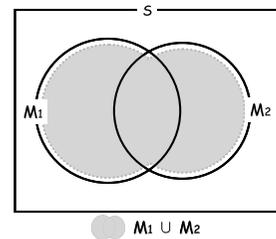


図6 委員会方式での上限

Fig. 6 Upper bound for committee-based parsing.

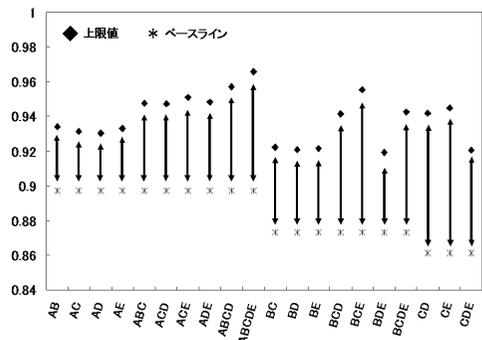


図7 各委員を組み合わせたとときの上限値とベースライン  
Fig. 7 Upper bound and base line.

受け正解率)をベースラインとする。委員会方式では、このベースラインを超える精度を得てはじめて精度向上が達成されたことになる。

各委員を組み合わせたとときの上限値(印)およびベースライン(\*印)を図7に示す。図より、どの委員の組合せにおいてもベースラインよりも上限値が上回っていることが分かる。このことは既存の複数の解析器を委員会方式の委員とすることで個々のどの解析器よりも精度が良くなる可能性があることを表している。

4.2 結果と考察

4.2.1 確率的投票の効果

まず、確率的投票の効果について調べる。図8および図9中の印は、確率的投票を行い、重み標準化戦略Simpleを用いた場合の結果を表しており、印は確率的投票を行わない場合を模擬するために重み標準化戦略No(式(10))を用いた場合の結果を表している。横軸は個々の解析器で作成した委員会を、図8の縦軸は総係り受け正解率、図9の縦軸は11-point正解率を表している。ここでは、確率的投票を行った場合に多重投票の効果が結果に反映されるのを防ぐために次式

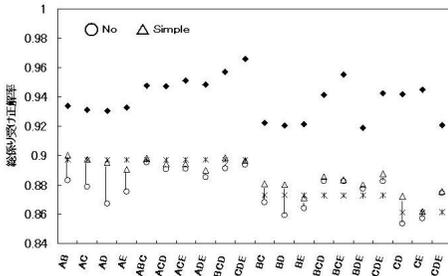


図 8 確率的投票の結果 (総係り受け正解率)

Fig. 8 Results: probabilistic voting (total accuracy).

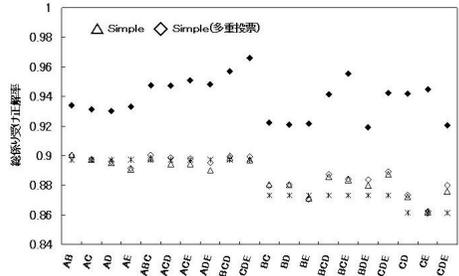


図 10 多重投票の結果 (総係り受け正解率)

Fig. 10 Results: multiple voting (total accuracy).

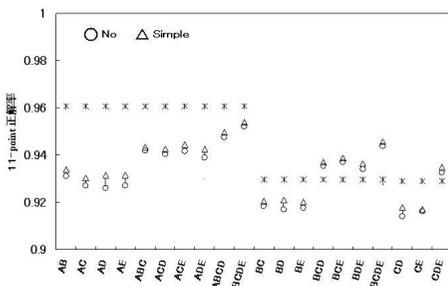


図 9 確率的投票の結果 (11-point 正解率)

Fig. 9 Results: probabilistic voting (11-point accuracy).

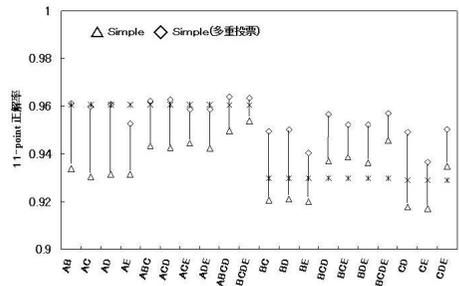


図 11 多重投票の結果 (11-point 正解率)

Fig. 11 Results: multiple voting (11-point accuracy).

$$w'_{ij} = \begin{cases} w_{ij} & (\text{if } j = \arg \max_k w_{ik}) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

によって 2 位以下の候補の重みを無効とした。また、参考のため上限値(印)とベースライン(\*印)も掲載する(以下、グラフ中の 印は上限値、\*印はベースラインを示す)。

No 重みなし。

$$w'_{ij} = \begin{cases} 1 & (\text{if } j = \arg \max_k w_{ik}) \\ 0 & (\text{otherwise}) \end{cases} \quad (10)$$

図 8, 図 9 より、いずれの委員会においても確率的投票を行わない場合に比べると確率的投票を行った方が解析精度が良いことが分かる。このことは、確率的投票が精度の向上に貢献できることを示唆している。しかしながら、ベースラインに精度が達していない委員会も多く見受けられる。特に、今回用いた解析器の中で最も性能の良い解析器 A (表 1 参照) が委員として含まれる場合はベースラインの値が高く設定されるために、委員会の精度は軒並みこの値を下回っている。またベースラインに達していたとしても上限値とは大きな開きがあることから、確率的投票だけでは不十分であることも分かる。

#### 4.2.2 多重投票の効果

次に、多重投票の効果について調べる。図 10, 図 11 中の 2 つのグラフは先ほどと同様、重み標準化 Simple を用いた結果を表している。印は式 (9) を適用して多重投票を無効にした場合の結果で、印は多重投票が有効な場合の結果である。総係り受け正解率では多重投票を行った場合と行わなかった場合の間にほとんど差が見られないが、11-point 正解率では顕著な差がでている。図 11 より、いずれの委員会においても多重投票を行った方が精度が向上しており、多重投票が精度の向上に貢献することが分かる。解析器 A が委員として含まれる場合にはベースラインに達していない委員会もあるが、その他の委員会ではほぼすべてにおいてベースラインを上回っており、確率的投票に加えて多重投票を行うことである程度の精度を保証することができると思われる。

また、2 つの解析器で委員会を構成した場合の誤り削減率がその他の場合に比べて比較的高く、5 つの解析器で構成した場合は低くなっている(表 3 参照)。これは、構成委員がある程度数そろえば 1 位の係り先のみで委員会としての係り先が決定され、2 位以下の係り先情報はそれほど影響を与えないためであると考え

表 3 図 11 の結果に対応する誤り削減率

Table 3 Error reduction rate.

委員会	誤り削減率	委員会	誤り削減率
AB	0.4139	BC	0.3640
AC	0.4232	BD	0.3684
AD	0.4263	BE	0.2525
AE	0.3080	BCD	0.3122
ABC	0.3339	BCE	0.2199
ACD	0.3484	BDE	0.2496
ACE	0.2590	BCDE	0.2114
ADE	0.2813	CD	0.3815
ABCD	0.2857	CE	0.2337
ABCDE	0.2104	CDE	0.2381

えられる。逆にいえば、構成委員が少ない場合は、1位の係り先のみでは情報が不十分であり、2位以下の係り先情報が精度向上に役立つことを示している。

4.2.3 重み標準化の効果

次に重み標準化の効果について調査する。重み標準化戦略 Normal, Class で用いる係り受け確率-係り受け正解率曲線を解析器ごとに獲得するため、実験データに対して 5 分割の交差検定を行った。

Class で用いる問題クラスについては、まず、問題の難易度（係り先の特定のしにくさ）、経験的な知見に基づき、格助詞を中心にして係り文節を 33 種類の問題クラスに分類した（付録参照）。このうち、出現頻度の少ない問題クラスでは係り受け確率-係り受け正解率曲線の獲得が困難であるなどの理由から、33 種の問題クラスすべてを用いることを避け、試験的にすべての解析器について以下に示す 14 の問題クラスを用いることとした

- [動詞連体] [動詞連用] [助詞 0, 動詞] [の 0, の]
- [が 0] [を 0] [に 0] [で 0] [と 0] [は 0]
- [は 1] [の 0] [助詞並列] 上記以外の残り

上記とは別に、「信頼性の高い係り受け確率 = 係り受け正解率の高い係り受け確率」と解釈し「信頼性の高い係り受け確率を持つ問題クラスは、他の問題クラスからの悪影響を防ぐために、独立な問題クラスとする」という基準を設けて入手で問題クラスを作成（[接続詞]、[の 0]、[指示詞 0]、[動詞体言] など）して実験を行ってみた。しかし予想していたほどの効果はなく、この問題クラスの基準はうまく作用しなかった。

また「他の問題クラスへの悪影響を防ぐために、信頼性の低い係り受け確率を持つ問題クラスは独立な問題クラスとする」という基準を設けて、経験的に係り先の特定が困難な問題クラス（[サ変名詞 1]、[副詞的名詞 1]、[時相名詞 1] など）を切り分けた実験も行った。この場合は全体に対して切り分けた問題クラスの割合が少ないために、精度への影響がほとんどでなかった。

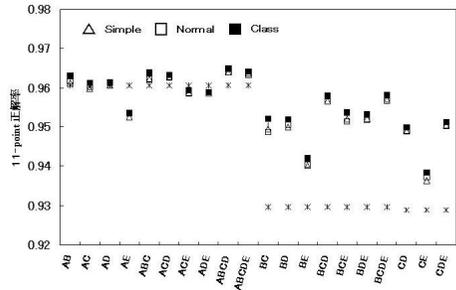


図 12 重み標準化の結果（11-point 正解率）

Fig. 12 Results: weighting standardization (11-point accuracy).

解析結果を図 12 に示す。重み標準化戦略 Normal, Class とともに、若干ではあるが重み標準化を行わない場合（Simple）に比べて精度が良い。このことは重み標準化が精度の向上に貢献することを示唆している。しかしながら、その差は有意なものではなく、その他の標準化戦略を用いて詳細な分析を追って行う必要がある。

最も精度の低い解析器 E を構成委員として追加する（たとえば AD から ADE）ことにより、精度の低下が見られる。問題クラスごとの精度を見てもその傾向は現れていた。たとえば問題クラス [の 0] の総係り受け正解率は 95.98% から 95.73% に下がっており、問題クラス [動詞連体] では 94.99% から 94.91% に低下していた。しかし、解析器 E が比較的良好な精度を持つ問題クラス [動詞連用] などでは逆に精度が良くなっていった（81.23% → 83.18%）。このことは、自動分類によって個々の解析器の特徴となる問題クラスをうまく分離できれば、精度の向上が見込める可能性があることを示唆している。

4.2.4 問題クラス分類の自動化に向けて（クロスエントロピーと精度の関係）

問題クラスの自動分類の際の指標として、確率言語モデルのクロス・エントロピーが利用できる可能性がある。この点について予備的な調査を行った。

テストセット  $S$  中の任意の入力文  $s$  のクロス・エントロピー  $h$ 、またテストセット全体でのクロス・エントロピー  $H$  は以下の式で計算される。

$$h(s) = -\frac{1}{|s|-2} \sum_{i=1}^{|s|-2} \sum_{j=i+1}^{|s|} \quad (11)$$

ここでは、11-point 正解率のみ掲載している。総係り受け正解率ではほとんど精度変化が生じていなかった。いずれの結果も多重投票を考慮した結果である。

表4 クロス・エントロピーと精度の関係  
Table 4 Cross entropy and accuracy.

委員会		Simple	Normal	Class
AB	$H_c(wf)$	0.0000	0.0538	0.3564
	$A_{11}$	0.9612	0.9616	0.9631
AC	$H_c(wf)$	0.0000	0.6862	0.2390
	$A_{11}$	0.9598	0.9604	0.9612
AD	$H_c(wf)$	0.0000	0.4277	0.7126
	$A_{11}$	0.9607	0.9613	0.9614
BC	$H_c(wf)$	0.0000	0.0853	0.0754
	$A_{11}$	0.9495	0.9488	0.9521
BD	$H_c(wf)$	0.0000	0.0318	0.1680
	$A_{11}$	0.9501	0.9506	0.9520
CD	$H_c(wf)$	0.0000	0.0295	0.0690
	$A_{11}$	0.9491	0.9492	0.9500

$A_{11}$  は 11-point 正解率の値を示す。

$$Q(r(s, i, j) | s) \log P(r(s, i, j) | s)$$

$$Q(r(s, i, j) | s) = \begin{cases} 1 & (\text{if } r(s, i, j) \text{ is correct}) \\ 0 & (\text{otherwise}) \end{cases}$$

$$H = \frac{1}{|S|} \sum_{s \in S} h(s) \tag{12}$$

$Q(r(s, i, j) | s)$  は 2 値関数であり、係り文節  $b_i$  の正しい係り先が  $b_j$  のときに 1 を返し、それ以外の場合は 0 を返す。|S| はテストセット中の文数を表し、|s| は文  $s$  の文節数を表す。

ここでは重み標準化戦略を変更した際の変化量を観察するために重み標準化戦略 Simple を基準にして、Simple との差分に注目する。つまり次式の  $H_c(wf)$  を評価値として算出することとした。

$$H_c(wf) = \sum_{M_k} (H(M_k, \text{Simple}) - H(M_k, wf))$$

$$wf = (\text{Simple} | \text{Normal} | \text{Class}) \tag{13}$$

ただし、 $H(M_k, wf)$  は、重み標準化戦略  $wf$  を用いた場合の解析器  $M_k$  によるクロス・エントロピー  $H$  を表す。

表 4 に構成委員数 2 の場合の計算結果を示す。表より、 $H_c(wf)$  と 11-point 正解率の間には有意な差ではないものの若干の相関関係がみられる。このことは重み標準化戦略 Class での問題クラス決定において、クロス・エントロピーを指標とすることにより、最適

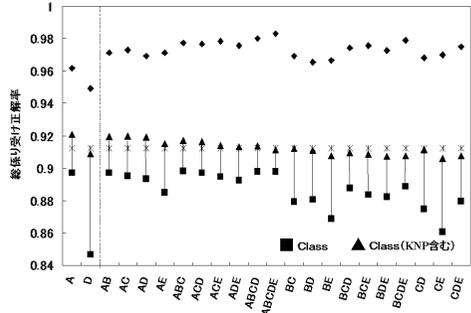


図 13 KNP を委員に加えた場合の精度 (総係り受け正解率)  
Fig. 13 Results: included KNP (total accuracy).

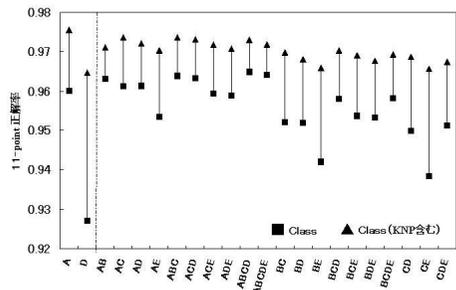


図 14 KNP を委員に加えた場合の精度 (11-point 正解率)  
Fig. 14 Results: included KNP (11-point accuracy).

な問題クラスが獲得できる可能性があることを示唆している。

4.2.5 非統計的な(ルールベースの)解析器の導入  
ここまで、統計ベースの解析器を用いて評価実験を行ってきたが、3.2 節で述べた枠組みでは、非統計的な(ルールベースの)解析器も委員として追加することができる。そこで、京大コーパスを解析対象とした解析精度において、これまでに報告されている係り受け解析器の中で最も性能の良い解析器である KNP<sup>17)</sup> を委員に加えた追実験を実施した。

実験に用いたデータは、ここまでの実験と同様の文節区切りを統一化した 14,234 文である。実験データによる KNP の総係り受け正解率は 91.24%であった。

KNP はルールベースの解析器であるため、係り受け確率行列の要素としては 0 か 1 の 2 値しか持たない。そこで、総係り受け正解率を票の重みと見なす重み標準化戦略 TA を用いて重みづけを行った。ただし、問題クラスを考慮して、それぞれの問題クラスごとの総係り受け正解率を票の重みとして用いている。TA 総係り受け正解率を利用する。

図 6 中の部分集合  $M_1 \cap M_2$  と  $\overline{M_1 \cup M_2}$  に含まれる係り文節については、精度変化に寄与する見込みがなく、係り受け確率を補正する(標準化)する必要性が希薄であるため、計算対象外として扱った。

$$w_{ij} = \begin{cases} TAC_{b_i} & (\text{if } j = KNP \text{ の出力した係り先}) \\ 0 & (\text{if } j \leq i) \\ \frac{1-TAC_{b_i}}{|s|-i-1} & (\text{otherwise}) \end{cases} \quad (14)$$

$TAC_{b_i}$  は問題クラス  $C_{b_i}$  の総係り受け正解率、 $|s|$  は入力文の文節数を表している。

結果を図 13, 図 14 に示す。比較対象として KNP を加えていない場合の結果も同時に示す。印が KNP を加えず重み標準化戦略 Class を用いた場合の結果で、印が 印のセットに対してさらに KNP を加えた場合の結果を表している。

まず、すべての委員会において KNP を加えると精度が飛躍的に向上した。解析器 A と KNP を組み合わせた場合が最も良い精度となっており、総係り受け正解率での評価が 92.09%, 11-point 正解率での評価が 97.56% であった。

図 13 より、精度は向上しているが、KNP を加えたことによって上限値、ベースラインもともに上昇していることが分かる。ベースラインを超えた委員会は解析器 A を含む一部の委員会のみであり、上限値との間にも大きな差がある。このことを考慮すると、解析精度をあげるにはさらなる検討が必要であると考えられる。

ここで上限値が大きく上がったことは委員会方式による検討を重ねることによって、今後、精度向上の可能性を示している。KNP は、現在の統計ベースでは扱っていない種々の言語的な知見がルールに考慮されており、このことが上限値の上昇につながったと考えられる。

問題クラス別に解析精度を調査したところ、統計ベースでは扱いにくく、KNP でも精度の低下が見られる問題クラス ([サ変名詞 1] や [副詞的名詞 1]) において興味深い結果が得られた。上記に該当する一部の問題クラスでは、その他の委員が KNP よりも精度が劣っているにもかかわらず、委員会の結果としては精度が上がっていた。たとえば、解析器 A と KNP の組合せでは、問題クラス [サ変名詞 1] の総係り受け正解率での精度変化が (解析器 A : 66.46%, KNP : 81.32% 委員会 : 82.39%) となっており、問題クラス [副詞的名詞 1] では (解析器 A : 76.82%, KNP : 80.94% 委員会 : 82.32%) となっていた。このことから、

KNP の精度が低い問題クラスに対して特に精度が良くなるようにチューニングされた解析器が存在していれば (たとえ KNP よりも総係り受け正解率が劣っていたとしても)、委員会としての結果はさらに良くなる可能性があることが分かる。

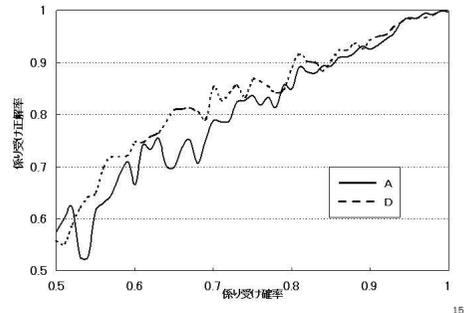


図 15 委員会の出力結果による P-A 曲線

Fig. 15 P-A curves included KNP.

図 15 に係り受け確率-係り受け正解率曲線を示す。委員会の構成委員の中にルールベースの解析器が入っていたとしても出力行列の要素はある程度妥当な値を持つことが分かる。解析器 D との組合せでは KNP 単独での精度を下回っているが、解析器 A との組合せでは KNP の精度を上回っている (図 13)。このことは、ルールベースの解析器とそれと同程度の性能を持つ統計ベースの解析器を構成委員とする委員会を作成すれば、ルールベースの解析器に対してその性能を落とさずに定量的な注釈が付与できることを意味する。

#### 4.3 関連研究との比較

ここでは、自然言語処理関連のタスクに委員会方式を適用した研究と提案した枠組みとの比較を行う。

パーズングに関する研究 Hendersonら<sup>12)</sup>は、Charniak モデル<sup>2)</sup>など、既存の代表的な統計的英文構文解析器 3 つを用いた解析実験から委員会方式の有効性を示している。Henderson らは、*parse hybridization* と *parser switching* という 2 種類の戦略を提案している。*parse hybridization* とは、個々の解析器が出力する構文構造の部分構造に対して多数決を行う手法であり、「委員となる解析器が満場一致する部分構造は最終出力に保存する」、「委員のだれからも支持されない部分構造を最終出力として作成すべきではない」という 2 つの制約を仮定する。一方、*parser switching* は、部分構造の情報を利用するが、最終的にはどれか 1 つの解析器 (の出力する構文構造) を選択する手法である。*parse hybridization* の具体的な処理としては、部分構造に対する重みなし多数決 (*constituent voting*) のほかに、naive Bayes 法より得られるスコアを用いてもっともらしい部分構造を選択する *naive bayes* 手法を提案している。また、*parser switching* では、独自に部分構造間の類似性尺度を定義し、その値が最も高い解析器 (の出力する構文構造) を選択す

る *similarity switching* 手法, *parse hybridization* のときと同様に naive Bayes 法より得られるスコアを用いて解析器 ( の出力する構文構造 ) を選択する *bayes switching* 手法をそれぞれ提案し, 評価まで行っている .

我々の枠組みに照らし合わせてみると, まず *parse hybridization* と *parser switching* はそれぞれ重み多数決過程における具体的な戦略であると見なせる . 特に, *parse hybridization* は, 入力文を係り受けごとに分割して多数決をとっている我々の実験と親和性が高い . また, naive Bayes 法を用いる手法は重み標準化を考慮した確率的投票に対応していると考えられる . 多重投票の概念は重みの決定処理の中で部分的に考慮されている可能性もあるが, 明示的には示されていない .

多義性解消に関する研究 Pedersen<sup>19)</sup> は, 2 段階の多数決過程を導入した委員会方式を多義性解消タスクに適用している . Pedersen の手法は, 語共起の情報を素性に用いた naive Bayes 法によってモデルを作成するが, その際, 多義語の前後に 81 種 ( 前後それぞれ 9 種ずつ ) の幅の異なる窓枠を設定し, 素性として考慮する語共起の文脈範囲を制限している . 作成された 81 個の個々のモデルは委員会での委員候補となるが, これらは窓枠幅の設定の仕方によってあらかじめ 9 つのカテゴリのいずれかに分類される .

処理の流れは, 多義語の入力に対して, まず 9 つのカテゴリからカテゴリの代表モデルをそれぞれ 1 つずつ選抜する ( 1 次多数決過程 ) . 代表モデル選抜の基準には, naive Bayes 法から得られるスコアを利用する . その後, 選抜された 9 つのモデルを委員とみても, 重みなしの多数決を行う ( 2 次多数決過程 ) .

この枠組みでは, 委員となるすべてのモデルは統一的に naive Bayes 法によりスコアを算出するため, 重み標準化戦略は仮定されていない . 2 次多数決過程において, naive Bayes 法から得られるスコアを重みとする重みつき多数決を採用した ( 我々の確率的投票に対応する ) 実験も同時に試みているが, 精度は逆に低下するという興味深い結果を報告している . また, Pedersen は多重投票に関する実験を行ってはいないが, Pedersen の実験では最終的に 9 つのモデルを構成委員としており, 構成委員数が比較的多い . 今回の我々の評価結果に基づく限りにおいては, 多重投票の適用による効果はあまり望めないと考えられる .

品詞タグづけに関する研究 van - Halteren<sup>23)</sup> は, 英文の品詞タグづけに委員会方式を適用している . この方式では, すべての要素が 1 あるいは 0 からなる入力行列について, 各モデル ( tagger ) が品詞タグの種類ごとの平均適合率・再現率から重みを付与し ( 我々の

重み標準化戦略 TA に対応 ), 重みつき多数決を行う .

Brill ら<sup>1)</sup> も英文の品詞タグづけに委員会方式を適用しており, Uni-gram モデル, N-gram モデル, Transformation-Based モデルおよび Maximum-Entropy モデルの 4 つのモデルで委員を構成して評価実験を行っている . 多数決戦略としては, 最も単純な重みなし多数決戦略のほか, *Contextual Cues* という概念に基づく 2 つの重みつき多数決戦略を提案している . まず, タグづけ対象語の前後に位置する 2 つの語およびその品詞を文脈 *Contextual Cues* と定義し, これらの情報に対して事例ベース学習を行う . そして, その学習結果に基づいてもっともらしい品詞を決定する手法とっともらしいモデル ( が出力する品詞 ) を決定する手法を採用している . Brill らの手法では, 事例ベース学習時に文脈情報を取り入れた確率的投票が扱われているが, ここでも多重投票は考慮されていない .

以上のことから提案した枠組みは, 確率的投票・多重投票により確率的に豊富な情報が意志決定に反映される点, また, 重み標準化と重み多数決の 2 つの過程を明示的に分離している点で, 既存の研究でとられている手法を一般化した枠組みであると考えられる .

## 5. おわりに

本稿では, 統計的部分解析手法と委員会方式という 2 つの概念を統合した枠組みについて考察した . 委員会方式の委員として統計的係り受け解析器を仮定すること, また, 各委員から得られる解析結果に基づき委員会での多数決により統計的部分解析を行うために, 従来の委員会方式の拡張を試みた .

従来の委員会方式からの拡張点は,

- ( 1 ) 係り受け確率を票の重みと見なして重みつきの票を投じる確率的投票,
  - ( 2 ) 委員間での票の重みの信頼性を標準化する重み標準化,
  - ( 3 ) 各係り文節に対して 2 位以下の係り先候補にも重みつきの票を投じる多重投票,
- の 3 点である .

以上の拡張は, 解析精度が向上することを理論的に保証するものではない . しかしながら, 評価実験より少なくとも以下のことが分かった .

- まず, 既存の統計的解析器の結果を組み合わせることにより, 個々の解析器よりも精度が良くなる可能性があることを確認した ( 図 7 ) .
- 委員の組合せによって精度変化には多少の揺れが

生じているものの、総合的には提案した枠組みが解析精度の向上に効果的であることを支持する結果を得た。

- 構成委員数が少ない場合には特に多重投票の効果が大きく、40%を越す誤り削減率が得られた。
- 重みの標準化では各解析器の特徴を問題クラスとして分離する戦略が精度の向上に有効に作用する見通しを得た。
- ルールベースの解析器 KNP を委員として加えた追実験を行ったところ、実験で用いたベンチマークによる評価に関する限り、総係り受け正解率で 92.09%と KNP ( 91.24% ) を上回る精度を得ることに成功した。

得られた評価実験の結果は委員会方式が有効であることを示唆していたが、あくまで 1 つの事例研究にすぎない。今後は、個々の問題クラスに対する委員の選択手法、あるいは、組合せ手法を考察していきたい。

謝辞 委員会方式については、奈良先端科学技術大学院大学の松本裕治氏に示唆に富む多くの助言をいただきました。同氏に深く感謝いたします。また、快く実験に協力して下さった NHK 放送技術研究所の江原暉将氏、奈良先端科学技術大学院大学(現在、日立中央研究所)の藤尾正和氏、東京大学(現在、IBM 東京基礎研究所)の金山博氏、東京工業大学(現在、北陸先端科学技術大学院大学)の白井清昭氏、通信総合研究所の内元清貴氏および京都大学(現在、東京大学)の黒橋禎夫氏の諸氏に深く感謝いたします。

### 参 考 文 献

- 1) Brill, E. and Wu, J.: Classifier Combination for Improved Lexical Disambiguation, *Proc. 17th COLING*, pp.191-195 (1998).
- 2) Charniak, E.: Statistical parsing with a context-free grammar and word statistics, *Proc. 15th National Conference on Artificial Intelligence* (1997).
- 3) Collins, M.J.: A new statistical parser based on bigram lexical dependencies, *Proc. 34th ACL*, pp.184-191 (1996).
- 4) Dagan, I. and Engelson, S.P.: Selective Sampling in Natural Language Learning, *Proc. IJCAI-95 Workshop on New Approches to Learning Natural Processing*, pp.41-48 (1995).
- 5) 江原暉将: 最大エントロピー法を用いた日本語文節間係り受け整合度の計算, 言語処理学会第 4 会年次大会予稿集, pp.382-385 (1998).
- 6) 藤尾正和, 松本裕治: 語の共起確率に基づく係り受け解析とその評価, 情報処理学会論文誌, Vol.40, No.12, pp.4201-4212 (1999).
- 7) Fiscus, J.G.: A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER), *EuroSpeech* (1997).
- 8) Frederking, R. and Nirenburg, S.: Three heads are better than one, *Proc. 4th ANLP* (1994).
- 9) Freund, Y.: Boosting a Weak Learning Algorithm by Majority, *Information and Computation*, Vol.121, pp.256-285 (1995).
- 10) Freund, Y. and Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, Vol.55, pp.119-139 (1997).
- 11) Haruno, M., Shirai, S. and Ooyama, Y.: Using Decision Trees to Construct a Practical Parser, *Machine Learning*, Vol.34, pp.131-149 (1999).
- 12) Henderson, J.C. and Brill, E.: Exploiting Diversity in Natural Language Processing: Combining Parsers, *Proc. 1999 Joint SIGDAT Conference on EMNLP and VLC*, pp.187-194 (1999).
- 13) 乾健太郎, 白井清昭, 田中穂積, 徳永健伸: 統計に基づく部分係り受け解析, 言語処理学会第 4 会年次大会予稿集, pp.386-389 (1998).
- 14) 乾 孝司, 乾健太郎: 統計的部分構文解析器のふるまいについて, 言語処理学会第 5 会年次大会「構文解析—現状の分析と今後の展望」ワークショップ論文集, pp.33-40 (1999).
- 15) Jensen, K., Heidorn, G.E. and Richardson, S.D.: *Natural language processing: The PLNLP Approach*, Kluwer Academic Publishers (1993).
- 16) Kanayama, H., Torisawa, K., Mitsuisi, Y. and Tsujii, J.: A Hybrid Japanese Parser with Hand-crafted Grammar and Statistics, *Proc. 18th COLING*, pp.411-417 (2000).
- 17) Kurohashi, S. and Nagao, M.: KN Parser: Japanese Dependency/Case Structure Analyzer, *Proc. International Workshop on Sharable Natural Language Resources*, pp.48-55 (1994).
- 18) Kurohashi, S. and Nagao, M.: Building a Japanese parsed corpus while improving the parsing system, *Proc. NLP/RS*, pp.151-156 (1997).
- 19) Pedersen, T.: A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation, *Proc. NAACL*, pp.63-69 (2000).
- 20) Schapire, R.: The Strength of Weak Learnability, *Machine Learning*, Vol.5, pp.197-227 (1990).
- 21) Shirai, K., Inui, K., Tokunaga, T. and Tanaka, H.: An empirical evaluation on statistical pars-

表5 文節の問題クラス  
Table 5 Problem classification.

1: [動詞連体 0, の] (提唱する地球規模の)	2: [動詞連体] (仕えた)
3: [動詞連用] (勝って)	4: [助詞 0, 動詞] (休暇を楽しむ)
5: [時相名詞 1] (今年, )	6: [時相名詞 0] (毎年)
7: [副詞的名詞 1] (結果, )	8: [副詞的名詞 0] (この時)
9: [数詞] (三十一日, )	10: [サ変名詞 1] (発達, )
11: [の 0, の] (大会関係者の心の)	12: [が 0] (世界が)
13: [が 1] (挑戦が, )	14: [を 0] (苗木を)
15: [を 1] (関係を, )	16: [に 0] (文章に)
17: [に 1] (対象に, )	18: [で 0] (家庭で)
19: [で 1] (サークルで, )	20: [と 0] (棋戦と)
21: [と 1] (技術と, )	22: [は 0] (漁業は)
23: [は 1] (漁業は, )	24: [の 0] (晴れ姿の)
25: [の 1] (国務省の, )	26: [助詞並列] (銀行や)
27: [形容詞連体 0] (近い)	28: [形容詞連用] (長く)
29: [指示詞 0] (その)	30: [副詞 1] (結局, )
31: [副詞 0] (互いに)	32: [接続詞] (そして)
33: [他] (雪)	

左側の四角括弧が問題クラスを表す。右側の丸括弧は該当する文節例である。

後続する文節を指定する問題クラス(No1:, 4:, 11:)については, 該当する文節部分に下線を引いている。

ing of Japanese sentences using a lexical association statistics, *Proc. 3rd EMNLP*, pp.80–87 (1998).

- 22) Uchimoto, K., Sekine, S. and Isahara, H.: Japanese dependency structure analysis based on maximum entropy models, *Proc. 13th EACL*, pp.196–203 (1999).
- 23) van Halteren, H., Zavrel, J. and Daelemans, W.: Improving data driven wordclass tagging by system combination, *Proc. 17th COLING* (1998).

#### 付録 実験で用いた問題クラス(表5)

“0”は文節内に読点が含まれていないこと, “1”は文節内に読点が含まれていることを表す。“,”印は後続する文節を指定している文節の問題クラスを示している。たとえば, クラス(No1:)[動詞連体 0, の]は動詞連体形で読点を含まない文節で, なおかつ助詞「の」を含んでいる文節が後続している問題クラスを表す。また, 重複する問題クラスに割り当て可能な文節は, より条件の厳しい問題クラスに割り当てることとした。たとえば, 文節「大会関係者の」は問題クラス[の 0, の]と[の 0]に該当する可能性がある。この場合は後続する文節に注目し, 後続文節が助詞「の」を含んで

いれば問題クラス[の 0, の], そうでなければ[の 0]に割り当てることとした。

(平成 13 年 5 月 30 日受付)

(平成 13 年 10 月 16 日採録)



乾 孝司

1999 年九州工業大学情報工学部卒業。2001 年同大学大学院情報工学研究科修士課程修了。同年, 奈良先端科学技術大学院大学情報科学研究科博士課程入学。現在に至る。自然言語処理の研究に従事。



乾 健太郎(正会員)

1990 年東京工業大学工学部卒業。1995 年同大学大学院情報理工学研究科博士課程修了。同研究科助手を経て, 1998 年より九州工業大学情報工学部助教授。現在に至る。1998 年より 2001 年まで科学技術振興事業団さきがけ研究 21 研究員を兼任。博士(工学)。自然言語処理の研究に従事。言語処理学会, 人工知能学会, 電子情報通信学会, ソフトウェア学会各会員。