

# 語義の曖昧性解消のための最適な属性選択

福 本 文 代†

本稿では、語義が付与されたコーパスから語義の曖昧性解消に必要な属性を抽出し、これを用いてテスト文中の多義を解消する手法を提案する。本手法では、語義の曖昧性解消に必要な属性は多義語が持つそれぞれの語義に依存して決まると仮定し、解消に最適な属性を多義語ごとに学習するのではなく、多義語が持つ語義ごとに学習するという方法を用いた。Leacock と Bruce らによって作成された 'line' と 'interest'、および SENSEVAL1 データを用いて実験を行った結果、'line' および 'interest' では、それぞれ 90.0%、および 90.1% の  $F$  値が得られ、SENSEVAL1 では、名詞、動詞、および形容詞それぞれ 79.5%、68.3%、および 73.3% のマイクロ平均  $F$  値が得られた。

## Toward Optimal Feature Selection for Word Sense Disambiguation

FUMIYO FUKUMOTO†

This paper describes a method for feature selection which is used for disambiguating word senses. In our method, sets of features which correspond to each different sense of an ambiguous word are selected by applying a machine learning technique. Empirical results which were tested on the two data, one is 'line' and 'interest' data, and another is SENSEVAL1 data, show that the performance of the method is comparable to the existing sense disambiguation techniques.

### 1. はじめに

語の曖昧性を解消する技術は、機械翻訳や自動抄録、あるいは情報検索システムなどの性能向上に必要な技術であることから、過去 40 年以上にわたり研究が行われてきた<sup>1)</sup>。特に 90 年代に入ってから、電子化された大量のコーパスが利用可能になったことを背景に、コーパスから統計的に学習した結果に基づき語の曖昧性を解消する研究が多く行われている<sup>2)~4)</sup>。

コーパスを利用した手法は、語義が付与されていないコーパスを用いる方法<sup>5),6)</sup>とあらかじめ人手により語義が付与されたコーパスを用いる手法<sup>7)~9)</sup>に分類できる。前者は、大量に入手可能なコーパスを学習データとして用いるという統計的処理にとって重要な利点を持っている反面、大規模な語彙に対する精度の面で検証の余地がある<sup>10)</sup>。一方、後者は精度が期待できる反面、コーパスへの語義付与に多大な労力とコストを要することが問題とされていた。このため、語義を付与すべきデータを選択的に抽出する手法なども提案されている<sup>11)</sup>。最近になり SENSEVAL<sup>12)</sup>や RWC<sup>13)</sup>など国内外において語義を付与した大量のコーパスが

整備されると同時に共通のコーパスにおけるシステム評価の場も用意されるようになったことから、語義付きコーパスを用いた研究が注目を集めている。

語義付きコーパスを利用した語の曖昧性解消手法は、従来より機械学習を用いた方法が多く提案されている。これは、 $k$ NN 法、決定木、Naive-Bayes 法などの機械学習を用いて解消に必要なモデルを導出し、それを用いてテスト文中の語義の曖昧性を解消する方法である<sup>2)~4),14)~17)</sup>。これらの手法の精度は属性選択、すなわちどのような属性を用いて多義語を含む文脈を表すかに依存する。たとえば、Yarowsky は、属性として共起関係に注目し、多義語と共起する名詞がその多義語の意味を 90~99% の高い精度 (accuracy) で一意に決めることが可能であることを実験により示している<sup>18)</sup>。

一方、多義語の意味は形態、構文、意味など複数の情報が相互に関係しそれらに依存して決まることから、高精度で解消を行うために多義語を含む文脈を複数の種類から成る属性で表現し、解消に役立てるという研究も多く提案されている<sup>16),19),20)</sup>。Towell らは多義語を含む文脈を topical と local と呼ばれる 2 つの属性を用いて表現している<sup>20)</sup>。topical 属性とは、多義語を含む文の 2 文以内にあるすべての周辺語であり、local 属性は、多義語の前後 3 語以内に位置する単語、

† 山梨大学工学部コンピュータ・メディア工学科  
Department of Computer Science and Media Engineering, Faculty of Engineering, Yamanashi University

および共起関係としている。Naive Bayesian classifier を適用し解消を行った結果、動詞 ‘serve’ および名詞 ‘line’ については、2つの属性を用いた方が、それぞれ単独で用いるよりも高い精度で解消できたことが報告されている。しかし、形容詞 ‘hard’ の場合には、local 属性のみを用いた結果と topical と local 属性の両方を用いた場合の精度がともに 83%であったことから、属性の種類を増やすことが必ずしも精度に貢献しないことを示唆している。このことは、Ng らの研究でも報告されている<sup>16)</sup>。彼らは、品詞情報のほか、派生情報、多義語の周辺語の情報、共起関係、および動詞と目的語の構文関係を用いて文脈を表現し、学習アルゴリズムとして Nearest-Neighbor classifier を適用し多義語の解消を行った。‘interest’ を用いた実験の結果、共起関係のみの精度は 80%であり、すべての属性を用いた精度は共起関係を属性として用いた場合と比較すると 7%向上しただけであった。原因の 1 つに、過学習の問題があげられる。これは、多義語が持つすべての語義に対して同一の属性(属性集合)を用いて解消しているためである。したがって、その属性集合を用いて特定の語義の曖昧性を解消するという点に注目した場合には、その解消に不必要な属性も含まれている可能性がある。また、高い解消精度を達成するために大量の属性を用いると、その中には解消に不必要な属性も含まれているため、結果的に精度を下げてしまう。

本稿では、この問題に対処するため、機械学習 SVM を用いて解消に必要な属性を語義ごとに学習し、これを用いてテスト文中の語義の曖昧性を解消する手法を提案する。本手法では、語義の曖昧性解消に必要な属性は多義語が持つ語義に依存して決まると仮定する。ここで、Bruce らによって作成された多義語 ‘interest’ を含む Wall Street Journal からの例文<sup>19)</sup>について考えてみる。

- (1) With uncertainty still hanging over ‘interest rates’ and the dollar, the market failed to find a focus that might lead to further investor commitments, they said.
  - (2) Mr. Johnson succeeds Harry W. Sherman, who resigned to pursue other ‘interests’, in both positions.
- (1) と (2) の ‘interest’ はそれぞれ、‘money paid for the use of money’、および ‘activity, subject, etc., which one gives time and attention to’ の語義を示す。‘interest’ が (1) における ‘interest’ の語義、すなわち ‘money paid for the use of money’ の語義を持つ

場合には名詞句 ‘interest rate’ として出現する機会が多い。実際、Bruce らによって作成された ‘interest’ のデータにおいて、2,369 文中 883 文に出現する ‘interest’ が名詞句 ‘interest rate(s)’ を構成し、それらはすべて、‘money paid for the use of money’ の意味であった。一方、‘interest’ が (2) における ‘interest’ の語義、すなわち ‘activity, subject, etc., which one gives time and attention to’ である場合には、動詞 ‘pursue’ と共起して出現することが多い。‘interest’ data において、‘interest’ と共起する ‘pursue’ は 29 存在し、それら ‘interest’ はすべて ‘activity, subject, etc.’ の意味であった。したがって、隣接語の属性は、(1) における ‘interest’ の語義の曖昧性を解消するのに有効である反面、(2) の ‘interest’ の隣接語は ‘other’ であり、これは ‘interest’ が (2) 以外の語義を持つ場合にも隣接語として用いられることから、(2) における ‘interest’ の語義の曖昧性を解消するのに有効な属性であるとはいえない。我々はこの性質を利用し、解消に必要な属性を多義語ごとに学習するのではなく、多義語が持つ語義ごとに分けて学習し、得られた属性を用いてテスト文中の語義の曖昧性を解消した。

本手法では機械学習として Support Vector Machine (SVM) を用いる。SVM は、 $n$  次元 Euclid 空間上に配置されたデータを 2 分するための学習アルゴリズムであり、高次元の入力属性を用いても過学習が起こりにくいことから、文書分類<sup>21)~23)</sup>や固有表現の自動抽出<sup>24)</sup>、あるいは未知語の品詞推定<sup>25)</sup>や Chunking<sup>26)</sup>などで広く利用されている。

本稿の構成は以下のとおりである。次章で属性による多義語の表現について説明する。3 章では機械学習である Support Vector Machine (SVM) の概略について述べ、SVM が大量の入力属性に対しても過学習を回避できる理論的根拠について説明する。4 章では、SVM を用いて語義ごとに最適な属性を学習する手法、およびその結果を用いた語義の曖昧性解消手法について述べる。5 章では実験について報告し、6 章で結論を述べる。

## 2. 属性による多義語の表現

多義語の精度は、多義語を含む文脈をどのような属性を用いて表現するかに依存する。本研究では、先行研究の中で、語の形態的な情報を属性として定義した Mooney<sup>27)</sup>、Leacock ら<sup>9)</sup>、Towell ら<sup>20)</sup>、および Pedersen<sup>28)</sup>の研究に基づき、8 種類の属性を定義した。多義語は、この属性を用いて表され、各語義ごとに必要な属性は学習の結果、この 8 つの属性の組合せ

表 1 属性とその略語

Table 1 Contextual features and their abbreviations.

	属性		
	単語の種類	多義語との位置関係	略語
1	多義語の周辺語と名詞句	左, 中央, あるいは右	np_inc_lcr
2	名詞句の部分単語列	なし	np_inc
3	多義語の周辺語と名詞句	左, 中央, あるいは右	np_lcr
4		なし	np
5	多義語の周辺語	左, あるいは右	w_lr
6		なし	w
7	動詞と名詞	左, あるいは右	cw_lr
8		なし	cw

表 2 多義語 ‘interest’ を含む文 (3) の属性 3 による表現

Table 2 Representation of context.

they/pp(l),receive/vb(l),minority/nn(l),interests/nns(c),in/in(r),the/dt(r) company/nn(r),minority/nn_interests/nns(c),the/dt_company/nn(r)
--

で表現される。属性は、多義語の周辺に位置する単語の種類と、その単語と多義語との位置関係を用いて表現される。

#### (1) 多義語の周辺に位置する単語の種類

我々は単語の種類として 4 種類、すなわち、(i) 多義語の周辺に位置する語と名詞句、および名詞句を構成する部分単語列、(ii) 多義語の周辺に位置する語と名詞句、(iii) 多義語の周辺に位置する語、(iv) 多義語の周辺に位置する動詞と名詞を用いた。これらはいずれも品詞の情報が付与されている。多義語の周辺に位置する語は、多義語を含む一文内に存在する語とした。また、(i) において、たとえば名詞句が ‘high interest rate’ の場合には、その部分単語列は、‘high interest’、および ‘interest rate’ とした。

#### (2) 位置関係

多義語の周辺に位置する語の位置情報として 3 種類、すなわち、(i) 周辺語が多義語の右側（後方）に位置する、(ii) 左側（前方）に位置する、(iii) 多義語と周辺語が同じ名詞句を構成する要素になっている場合には中央に位置するという情報を用いた。

表 1 に本研究で使用した 8 つの属性とその略語を示す。

表 1 において、‘多義語との位置関係’における‘なし’は、多義語の周辺に位置する語に対して位置に関する情報が付与されていないことを示す。たとえば略語

‘cw’は、属性となる語が名詞あるいは動詞であり、その位置情報は付与されていないことを示す。今、文 (3) を多義語 ‘interest’ を含む文とする。

(3) They/pp receive/vb [ minority/nn interests/nns ] in/in [ the/dt company/nn ] ./.

文 (3) において ‘x/y’ の ‘x’ と ‘y’ はそれぞれ単語とその品詞を示す。[ ] で囲まれた単語列は名詞句を示す。文 (3) を表 1 における 3 の属性を用いて表現すると表 2 となる。

表 2 の (l), (r), および (c) は語の位置情報を示し、その語が多義語の左側か右側、あるいは多義語と同じ名詞句を構成する要素になっていることを示す。表 2 において、各要素の値は、多義語を含む文中におけるその要素（語）の出現回数とし、文中に出現しない場合にはゼロとした。

### 3. SVM

Support Vector Machine (SVM) は、Vapnik により提案された二値分類のための学習アルゴリズムであり<sup>29)</sup>、 $n$  次元 Euclid 空間上に配置されたデータを正例集合と負例集合に分離する超平面 (hyperplane) を得る学習手法である。今、属性で表現された訓練データの集合 (多義語の集合) を  $n$  次元 Euclid 空間上の hyperplane

$$\vec{w} \cdot \vec{x} + b = 0 \quad (1)$$

で分離する問題を考える。図 1 に示すように、近接する正例と負例のデータ間の間隔が大きいほど、精度良くテストデータを分類できる。

ここで、 $\vec{x}$  を任意のデータ点 ( $\vec{x} \in X$ ) とし、 $\vec{w}$  と  $b$  は訓練データが線形で分離できる場合に得られる値

実験において、‘line’ と SENSEVAL1 データの名詞は、1 つの多義語に対し多義語を含む 2 文以内にある語を用いた。

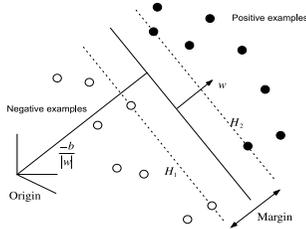


図1 線形の場合における分離超平面

Fig. 1 The decision surface of a linear SVM.

とする． $D = \{(y_i, \vec{x}_i)\} (1 \leq i \leq l)$  は訓練データの集合とし， $y_i \in \{\pm 1\}$  はデータ  $\vec{x}$  が正例 (+1) か負例 (-1) かを表すスカラとする．今，式 (2) と (3) で示される 2 つの hyperplane を考える．

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \quad \text{for } y_i = +1 \quad (2)$$

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \quad \text{for } y_i = -1 \quad (3)$$

式 (2)，および (3) は separating hyperplane とよばれ，まとめて書くと，

$$y_i [w \cdot x_i + b] \geq 1, \quad i = 1, \dots, l \quad (4)$$

となる．ここで，式 (2) を満たす点を考える．この点と式 (2) との距離は， $\frac{|1-b|}{\|\vec{w}\|}$  となる．同様に，式 (3) を満たす点と式 (3) との距離は  $\frac{|-1-b|}{\|\vec{w}\|}$  となる． $d_+$ ，および  $d_-$  をそれぞれ，式 (2) と (3) で示される separating hyperplane から最も近い正例と負例の点までの距離とすると， $d_+ = d_- = \frac{1}{\|\vec{w}\|}$  となり，正負例間のマージンは  $\frac{2}{\|\vec{w}\|}$  となる．したがって，SVM は訓練データを正例と負例に分け，正負例間のマージンが最大，すなわち式 (2) と (3) において  $\|\vec{w}\|$  が最小になるような  $\vec{w}$  と  $b$  を求める問題に帰着できる．

この問題は Lagrange 乗数を導入し，2 次最適化問題として扱うことができる．具体的には，式 (5) が最小となるような係数  $\alpha_i$  の集合を求める最適化を行い，式 (6) の  $\alpha_i$  から  $w$  を求めることで，マージンが最大となるような超平面を得ることができる．

$$\begin{aligned} & - \sum_{i=1}^n \alpha_i + \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \text{s.t. } & \sum_{i=1}^n \alpha_i y_i = 0, \quad \forall i, \alpha_i \geq 0 \end{aligned} \quad (5)$$

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (6)$$

SVM は，最小の汎化誤差，すなわち訓練データにおける誤差と仮説空間の複雑さを抑える仮説を求める構造化リスク最小化に基づく手法である．

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2l/h)+1) - \log(\eta/4)}{l}} \quad (7)$$

$$h \leq \min \left( \frac{D^2}{M^2}, n \right) + 1 \quad (8)$$

式 (7) は，仮説  $\alpha$  の汎化誤差  $R(\alpha)$  が  $1 - \eta$  の確率で，右辺第 1 項で示される訓練データにおける仮説  $\alpha$  のエラー率  $R_{emp}(\alpha)$  と第 2 項で示される仮説  $\alpha$  の複雑さの和で抑えられることを示す．ここで  $l$  と  $h$  はそれぞれ訓練データの個数と VC 次元と呼ばれる仮説空間の複雑さを表す指標である．VC 次元は， $l$  個の訓練データを覆う直径  $D$  の球と，マージン  $M = \frac{2}{\|\vec{w}\|}$ ，および入力ベクトルの次元  $n$  により，式 (8) で示される上限値が存在することが証明されている<sup>29)</sup>．式 (8) において， $n$  が十分に大きい場合，VC 次元は  $\|\vec{w}\|$  に依存する．したがって，結果的に高次元の入力属性を用いても高い汎化能力を持つことが分かる．

SVM は任意のデータ点  $\vec{x}$  を高次元空間に非線型的に射影し，高次元空間上で hyperplane を構成することができるため，非線形なベクトル空間に対する学習も可能であり，結果的に複雑な問題でも精度良く分類できるといわれている．この射影関数を Kernel 関数とよぶ．非線形の問題を扱うためには，式 (5) 中の内積を Kernel 関数で置き換えることにより実現する．Kernel 関数には多くの種類がある<sup>29)</sup>．本稿における実験では線形のほか，属性の共起関係の有効性を検証するために式 (9) で示される  $d$  次の多項式関数において  $d = 1$  と  $d = 2$  を用いる．

$$K(\vec{x}, \vec{z}) = (\vec{x} \cdot \vec{z} + 1)^d \quad (9)$$

今，式 (9) において  $d$  が 2 次の場合に共起関係が扱えることを示す．式 (9) において  $d$  の値を 2 とし，訓練データの集合を  $\vec{x} = (x_1, x_2)$ ，および  $\vec{z} = (z_1, z_2)$  とすると 2 次元上のベクトル  $\vec{a} = (a_1, a_2)$  を以下で示す 6 次元に射影する関数  $\Phi(a)$  を得ることができる．

$$(a_1, a_2) \mapsto (a_1^2, a_2^2, \sqrt{2}a_1a_2, \sqrt{2}a_1, \sqrt{2}a_2, 1) \quad (10)$$

式 (10) は， $a_1$  と  $a_2$  の共起が属性として定義されていることを示す．式 (9) は関数  $\Phi$  を用いて以下のように表される．

$$\begin{aligned} K(\vec{x}, \vec{z}) &= \Phi(\vec{x}) \cdot \Phi(\vec{z}) \\ &= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1 x_2 z_1 z_2 + 2x_1 z_1 + 2x_2 z_2 + 1 \end{aligned} \quad (11)$$

一般に， $d \geq 2$  は  $d$  個の属性が共起することを示す<sup>30)</sup>．

## 4. 属性選択と語義の曖昧性解消

### 4.1 属性選択

多義語を持つ語義ごとに最適な属性を選択するために，訓練データと疑似テスト (held-out test) を用い

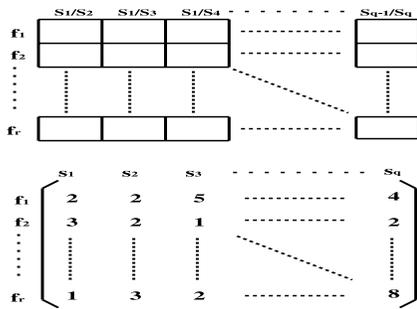


図2  $qC_2 \times r$  個からなる分類器と分類器による判定結果  
Fig. 2  $qC_2 \times r$  classifiers and disambiguation result.

る．まず多義語を含む訓練データと疑似テスト中の各文をベクトルで表現する． $x_i$  を多義語  $w_p$  を含む文とすると， $x_i$  の属性によるベクトル表現は式 (12) で示される．

$$\vec{x}_i = (w_1, \dots, w_n) \quad (12)$$

式 (12) において  $n$  は訓練データにおいて属性を用いて抽出された総異なり単語数を示す． $w_j$  ( $1 \leq j \leq n$ ) は， $w_j$  が  $x_i$  内に存在するときの頻度を示し，存在しない場合にはゼロとする．また， $x_i$  内の  $w_p$  が正例である場合に  $x_i$  は  $+1$  とし，負例である場合には  $-1$  とする．次にベクトルで表された訓練データに対して SVM を適用し，式 (2) および (3) で示される  $\vec{w}$  と  $b$  を求める．

SVM は任意のクラスに属するか否かを判定する二値分類のための学習アルゴリズムである．しかし，一般に多義語は 3 つ以上の語義を持つこともあるため，そのままでは扱うことができない．SVM を多値問題へ拡張する方法はいくつか提案されている<sup>(31),(32)</sup>．本稿では *pairwise* とよばれる方法を用いて学習，分類を行う<sup>(33)</sup>．*pairwise* 法は，多義語が  $q$  個の語義を持つ場合，任意の 2 つの語義に分離するような二値分類器を合計  $qC_2$  個作成し，テストデータの多義語は，各二値分類器で判定された語義のうち，個数が最も多い語義に判定する方法である．*pairwise* 法を用いた属性選択を以下に示す．

(1)  $w_p$  を  $q$  個の語義を持つ多義語とする．また， $r$  を属性のすべての組合せの総数とする．我々は  $qC_2 \times r$  個の訓練データに対して SVM を適用し，それぞれにおいて  $\vec{w}$  と  $b$  を求めた．その結果， $qC_2 \times r$  個の分類器が作成される．図 2 の上図に作成された分類器を示す．

図 2 の上図において，ボックスは分類器を示す．各分類器は，任意の 2 つの語義を判定する 2 値分類である．たとえば  $f_1$  と  $s_1/s_2$  で示された

四角は， $f_1$  の属性で作成された分類器を示し， $w_p$  が  $s_1$  の語義か  $s_2$  の語義であるかを判定する分類器である．この分類器は，訓練データ中の文に含まれる  $w_p$  が  $s_1$  の語義であるデータと  $s_2$  の語義であるデータに対して SVM を適用することにより得られる．

- (2) 各属性集合  $f_1, \dots, f_r$  に対して，疑似テスト文中の多義語  $w_p$  の語義を決定する．疑似テスト文中の  $w_p$  は， $qC_2$  個から成る各分類器で判定された語義の個数が最も多い語義  $s_x$  とする．
- (3)  $r$  個から成る属性集合ごとに，正しく判定された疑似テスト文の個数を求める．図 2 の下図の各値は正しく判定された疑似テスト文の個数を示す．
- (4) 語義ごとに，正しく判定された疑似テストの個数が最も多い属性集合を求める．たとえば，図 2 の下図において， $s_2$  の語義であるか否かを判定するのに最適な属性は， $f_r$  における疑似テストの正解数が 3 であり，属性集合  $f_1, f_2, \dots, f_r$  の中で最大であることから  $f_r$  となる．

#### 4.2 語義の曖昧性解消

抽出した属性集合を用いた語義の曖昧性解消手法を以下に示す．

- (1) 語義ごとに抽出した属性を用いて訓練データを表し，SVM を適用する．その結果， $qC_2 \times q$  個 の分類器が作成される．
- (2) テスト文中の  $w_p$  の語義は，語義  $s_x$  を判定するのに抽出された属性を用いて解消した結果， $s_x$  であると判定されたときに語義  $s_x$  であるとする．

## 5. 実 験

本手法の有効性を検証するために，(i) Leacock と Bruce らによって作成された ‘line’<sup>9)</sup> と ‘interest’<sup>19)</sup>，および (ii) SENSEVAL1<sup>34)</sup> の 2 種類のデータを用いて実験を行った．評価は以下の 6 項目について行った．

- 本手法の精度
- 語義数の違いによる精度への影響
- 訓練データ数の違いによる精度への影響
- 属性選択の有効性
- Kernel 関数の精度への影響
- 関連研究との比較

疑似テストの正解数が最大となる属性集合が複数ある場合には，それらすべてを抽出した．

1 語義あたり抽出された属性集合が複数ある場合には，その個数分の分類器が作成される．



表6 SENSEVAL1 data の多義解消結果  
Table 6 *F*-measure (SENSEVAL1 data).

名詞								
語	疑似テスト	テスト	語	疑似テスト	テスト	語	疑似テスト	テスト
accident	87.2	84.6	behaviour	97.0	95.6	bet	71.0	71.8
excess	80.1	79.0	float	93.4	93.3	giant	91.5	90.3
knee	55.0	54.9	onion	100.0	98.0	sack	83.2	83.8
scrap	40.3	36.6	shirt	90.2	88.7			
マイクロ平均 <i>F</i> 値 (疑似テスト/テスト) (80.1/79.5)								
動詞								
語	疑似テスト	テスト	語	疑似テスト	テスト	語	疑似テスト	テスト
amaze	99.0	98.5	bet	68.0	65.8	bother	80.0	80.4
bury	73.4	70.3	calculate	77.2	77.0	consume	55.6	54.3
derive	43.2	40.7	float	65.0	63.1	invade	52.0	50.2
promise	78.3	75.8	sack	84.0	83.7	scrap	94.0	94.0
seize	59.8	59.6						
マイクロ平均 <i>F</i> 値 (疑似テスト/テスト) (71.2/68.3)								
形容詞								
語	疑似テスト	テスト	語	疑似テスト	テスト	語	疑似テスト	テスト
brilliant	83.7	83.7	floating	58.3	59.3	generous	34.1	34.8
giant	74.3	72.2	modest	82.9	82.2	slight	75.5	72.5
wooden	96.0	94.3						
マイクロ平均 <i>F</i> 値 (疑似テスト/テスト) (75.0/73.3)								

表7 抽出した属性集合 ('line')  
Table 7 Extracted feature sets using 'line' (held-out test) data.

line1		line2		line3		line4		line5		line6	
属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値
w_lr	94.3	np	77.4	np_inc	80.5	np_inc_lcr	79.9	np_inc	96.9	np_inc_lcr	71.0
w		np_inc_lcr		np		np_inc		np		np_inc	
cw		np_inc				w_lr		w		cw_lr	
						w		cw		cw	

表8 抽出した属性集合 ('interest')  
Table 8 Extracted feature sets using 'interest' (held-out test) data.

interest1		interest2		interest3		interest4		interest5		interest6	
属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値	属性集合	<i>F</i> 値
np_lcr	93.4	248 sets of features	0	np_lcr	60.0	np_inc_lcr	62.8	np_inc_lcr	83.1	np_inc_lcr	99.5
w_lr				np		np_lcr		np_lcr		cw_lr	
w				w		w_lr		w_lr		cw	
								w			

テストのマイクロ平均 *F* 値を示す。表5において、'line' と 'interest' では、それぞれ 90.0%, 90.1% の *F* 値が得られた。また表6より SENSEVAL1 のテストデータでは、名詞、動詞、および形容詞それぞれ 79.5%, 68.3%, および 73.3% のマイクロ平均 *F* 値が得られた。

表5より、'line' と 'interest' はともに疑似テストとテストデータの *F* 値に差がないことが分かる。一方、SENSEVAL1 では名詞、動詞、および形容詞のいずれの場合にもテストデータのマイクロ平均 *F* 値は疑似テストよりも若干低下し、精度差はそれぞれ 0.6%, 2.9%, 1.7% であった。

本手法の属性選択は、語義ごとに正しく判定された疑似テストの個数が最大の属性集合のみ選択している。一方、Pedersen の手法は、数多くの精度の低い分類器を組み合わせて非常に高い精度を持つ分類器を実現するという方法であり、'line' と 'interest' を用いた実験の結果、テストデータの精度は疑似テストに比べ、それぞれ 3%, および 4% 向上している<sup>28)</sup>。今後は、尤度などを考慮に入れた属性選択の手法について検討する必要がある。

表7と表8は、それぞれ 'line' と 'interest' のデータを用いて抽出された属性集合を示す。

表7の 'line1' ~ 'line6'、および表8の 'interest1' ~

表9 ‘line’ および ‘interest’ の語義

Table 9 Distribution of senses for ‘line’ and ‘interest’.

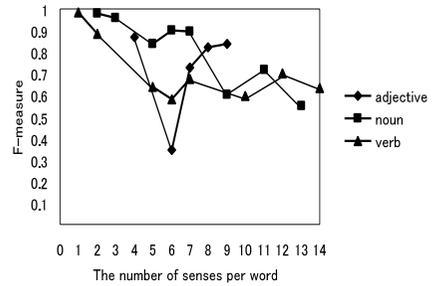
id	‘line’ の語義	訓練データ
line1	a thin, flexible object; cord	373
line2	formation of people or things; queue	349
line3	an artificial division; boundary	376
line4	telephone connection	429
line5	product	2,218
line6	written or spoken text	404
Total		4,149
id	‘interest’ の語義	訓練データ
interest1	readiness to give attention	361
interest2	causing attention to be given to	11
interest3	activity that one gives attention to	66
interest4	advantage, advancement or favor	178
interest5	a share in a company or business	500
interest6	money paid for the use of money	1,253
Total		2,369

‘interest6’ で示される各語義と訓練データ数を表9に示す。

表7と表8において各語義の1列目は、抽出された属性集合を示し、2列目は、疑似テストの $F$ 値を示す。表7と表8によると‘interest2’を除く‘line’と‘interest’のいずれの語義の場合にも、2~4種類の属性を用いた場合に最良の $F$ 値が得られていることが分かる。また、SENSEVAL1データの場合には $F$ 値が0%を除き、いずれも2~5種類の属性が抽出されていた。

表7において、たとえば、‘line1’の意味を示す‘an artificial division; boundary’の場合、抽出された属性は、‘左右の位置情報付きの周辺語’、‘位置情報なしの周辺語’、および‘動詞+名詞’であり、訓練データ中の‘line1’は動詞‘cut’、あるいは‘hang’と共起し、高頻度で出現していた。同様に、‘line3’の意味である‘a thin, flexible object; cord’の場合には、‘位置情報なしの周辺語’、名詞句および名詞句の部分単語列’、および‘位置情報なしの周辺語と名詞句’が属性として抽出されており、訓練データ中の‘line3’は‘draw the line between 名詞句’の形式で頻出していることから、各意味を特徴付ける属性が選択されていることが分かる。

表8によると‘interest2’の語義の $F$ 値はすべての属性の組合せのいずれの場合においても0%であった。原因として、表1で示される8つの属性の中に、多義解消に必要な属性が含まれていなかったことがあげられる。今回使用した‘interest2’の訓練データ(11文)中、8文に現れる‘interest’はいずれも、‘of interest’、あるいは‘of +形容詞+ interest’の形で現れた。一方、定義した8つの属性にはこれらの定型表現が反映され

図3 語義数と $F$ 値Fig. 3 # of senses per word and  $F$ -measure.

ていないため、結果的に正しく多義を解消することができなかったと考えられる。また、訓練データ数が少ないことも大きな要因として考えられる。今後は、使用する属性について再考すると同時に、訓練データとテストデータに対して統計手法を用いた単語の類似クラスを求めた結果を用いたり、WordNetの同義語クラスであるSynsetの情報を利用することで訓練データ数が少ない語義に対しても高精度で解消できるよう検討する必要がある。

### 5.2.2 語義数の違いによる精度への影響

図3は、多義語を語義数で分け、それぞれのマイクロ平均 $F$ 値を品詞ごとにプロットしたものである。

図3より、いずれの品詞の場合にも語義数が4以下の場合は、87%を超える精度が得られている。動詞や形容詞は語義数が5以上になるとマイクロ平均 $F$ 値が減少している。一方、名詞は語義数が7までは、84%以上の $F$ 値が得られていることから、語義数の違いによる精度への影響が少ないことが分かる。これは、名詞は動詞や形容詞と比べると、各語義のいずれもが多義語の周辺に位置する語の情報を用いて決定できる場合が多いためではないかと推測される。このことから、本研究で定義した属性は名詞に対してより効果があったといえる。

一方、形容詞は語義数が6のときに極端に $F$ 値が低下している。原因として、語義数が6である多義語は‘generous’であり、(i)各語義が持つ訓練データの個数はいずれも100文以下であったこと、(ii)6つの語義のいずれもが‘generous+名詞’と‘be動詞+generous’で出現するため、多義語の周辺語に基づく属性からは、各語義の違いを反映させるような特徴を抽出することができなかったことが考えられる。

### 5.2.3 訓練データ数の違いによる精度への影響

実験で使用したデータは、それぞれ語義数と各語義の訓練データ数が異なる。そこで、多義語の各語義が持つ訓練データ数のばらつき(分散値)が類似してい

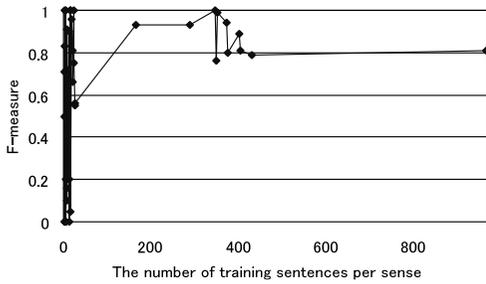


図4 語義あたりの訓練データ数と  $F$  値 (分散値小)

Fig. 4 # of training sentences per sense and  $F$ -measure.

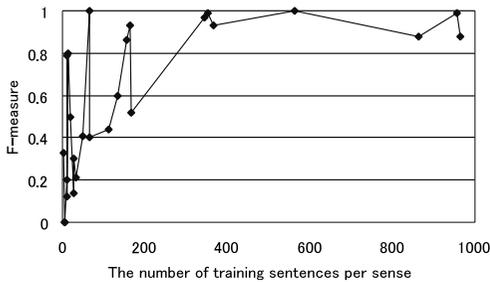


図5 語義あたりの訓練データ数と  $F$  値 (分散値大)

Fig. 5 # of training sentences per sense and  $F$ -measure.

る多義語どうしをグループ分けし、グループ内で1語義あたりの訓練データ数とその語義の  $F$  値との関係を検証した。図4は、分散値の低い11語(名詞である‘bet’, ‘float’, ‘onion’, ‘scrap’, ‘giant’, ‘interest’, ‘line’, および動詞である‘bet’, ‘consume’, ‘invade’, ‘promise’)の結果を示し、図5は分散値の高い7語(名詞である‘accident’, ‘behaviour’, ‘giant’, および動詞である‘promise’, ‘sack’, そして形容詞である‘giant’, ‘wooden’)の結果を示す。ただし、分散値が低い多義語のグループはいずれも1語義あたりの訓練データ数が30文以下であった。そこで、多義語が持つ語義のうち、1語義あたり150文以下である語義は排除し、かつ分散値が低くなるような多義語を新たに作成し、グループに加えた。分散値が低い11語中、‘giant’, ‘line’, ‘interest’, および‘promise’は、新たに作成した多義語である。たとえば6つの語義を持つ‘giant’から新たに作成した *giant\** は、2つの語義から成る多義語である。

図4によると1語義あたりの訓練データ数が30文以下の場合には、訓練データ数と  $F$  値に相関がみられない。一方、訓練データ数が352文の語義の場合に

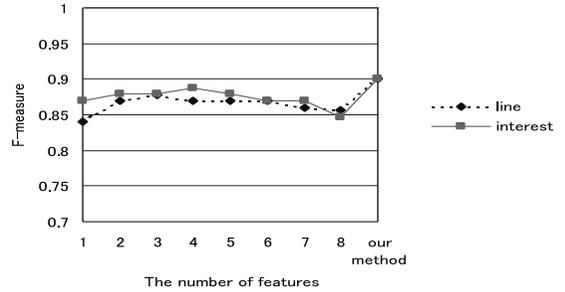


図6 ‘line’ と ‘interest’ における属性数と  $F$  値

Fig. 6 # of features and  $F$ -measure (‘line’ and ‘interest’).

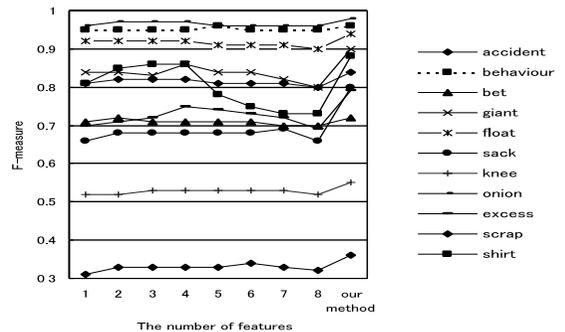


図7 SENSEVAL1 (名詞) における属性数と  $F$  値

Fig. 7 # of features and  $F$ -measure (SENSEVAL1, noun).

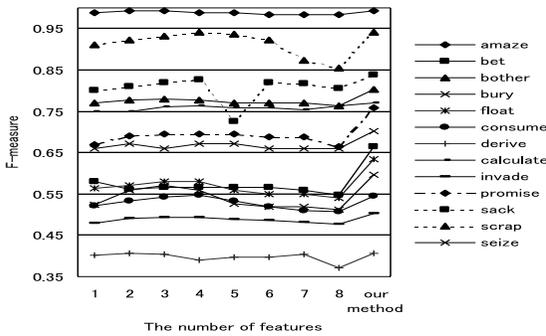
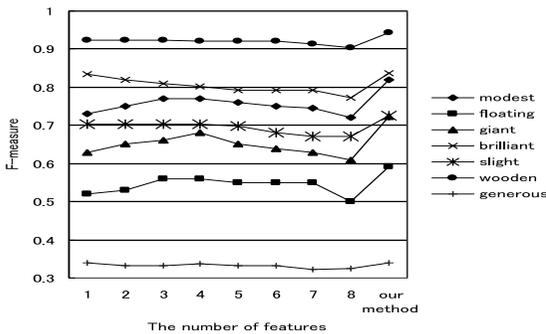
99%に達し、それ以上のデータ数のいずれの場合にも  $F$  値は80%を超えている。このことは、各語義が持つ訓練データ数のばらつきが大きいグループでもみられた。すなわち、図5において、1語義あたりの訓練データ数が200文以下の場合には両者に相関がみられず、訓練データ数が345文を超える語義の  $F$  値はいずれも88%を超えた。以上のことから1語義あたり350文以上の訓練データ数を持つ語義については、本手法は高い精度で解消できるといえる。

#### 5.2.4 属性選択の有効性

属性選択の有効性について検証するため、属性数の違いによる精度への影響を考察する。図6は、‘line’ と ‘interest’ における属性数と  $F$  値との関係を示し、図7、図8および図9はそれぞれ SENSEVAL1 の名詞、動詞、および形容詞における属性数と  $F$  値との関係を示す。各図における‘1’, ‘2’, ..., ‘8’は、それぞれ属性の組合せが  ${}_8C_{1,8}, C_2, \dots, {}_8C_8$  の中で最も高い精度が得られた属性集合の  $F$  値を示す。‘our method’ は本手法の  $F$  値を示す。

図6から図9で示されるいずれの多義語の場合においても本手法の  $F$  値が各属性の組合せの中で最も高いことが分かる。属性数と  $F$  値との関係に注

グループ内での多義語の個数を増やすため、品詞による語の区別は行わなかった。

図8 SENSEVAL1 (動詞)における属性数と  $F$  値Fig. 8 # of features and  $F$ -measure (SENSEVAL1, verb).図9 SENSEVAL1 (形容詞)における属性数と  $F$  値Fig. 9 # of features and  $F$ -measure (SENSEVAL1, adjectives).

目すると、SENSEVAL1の動詞‘bet’と形容詞‘brilliant’を除くすべての語において、1種類の属性を用いた場合よりも、複数の属性を用いた場合の方が良い精度が得られている。しかし、たとえば‘line’の場合、最も高い精度は3種類(‘np\_inc\_lcr’, ‘np\_lcr’, ‘cw’)で  $F$  値は87.7%、また、‘interest’の場合には4種類(‘np\_inc\_lcr’, ‘np\_lcr’, ‘w\_lr’, ‘w’)で  $F$  値は88.8%であり、ともにすべての属性を用いた場合でなかったことから、属性の組合せの個数は、必ずしも精度に比例していないことが分かる。また、一般に大量の属性を用いると過学習が起こりやすくなるが、属性数の増加にとまぬ必ずしも精度が低下していないことから、SVMが高次元の入力属性に対しても高い汎化能力を持つことが分かる。

‘line’と‘interest’における本手法と8つのすべての属性を用いた場合の精度差は、それぞれ4.3%(90.0–85.7)、および5.4%(90.1–84.7)であったことから、NgやPedersenらと同様、‘line’と‘interest’の場合には両者に大きな差はみられなかった。これは、定義した8つの属性の中には各語義の曖昧性解消に必要な属性が含まれ、かつ解消に不必要な属性が少な

表10 多項式関数の次元と  $F$  値(‘line’および‘interest’)Table 10  $F$ -measure using the  $d$ -th polynomial functions (‘line’ and ‘interest’).

語	$d = 1/d = 2$	語	$d = 1/d = 2$
line	90.0/90.2	interest	90.1/91.3

かったためであると考えられる。

一方、SENSEVAL1において本手法の  $F$  値がすべての属性を用いた場合の  $F$  値よりも10%以上高い精度が得られた多義語は、名詞と動詞、および形容詞でそれぞれ1語、3語、および3語存在し、‘excess’, ‘bet’, ‘scrap’, ‘promise’, ‘floating’, ‘giant’, および ‘modest’であった。属性選択は、あらかじめ定義された属性のうち、特定の属性集合に特定の語義の曖昧性を解消する属性が含まれているときに有効である。本手法で定義した8つの属性を用いた属性選択は、名詞、形容詞、および動詞のうち、狭いウィンドウサイズで語義が決まる動詞と形容詞においてより効果的に働いたといえる。

### 5.2.5 Kernel 関数の精度への影響

本手法で用いた Kernel 関数である多項式の次元について考察する。表10、および表11において  $d = 1$ 、および  $d = 2$  は、多項式が1次、および2次の結果を示す。多項式が2次のときは、属性となる2つの語のすべての組合せを共起関係として扱うことができる。1次と2次の  $F$  値を比較すると、両者に顕著な差が現れず、 $d = 2$  の場合にわずかに  $F$  値が向上した語は、‘line’, ‘interest’, および SENSEVAL1 データでは、‘accident’, ‘behaviour’, ‘sack’, および ‘wooden’のみであった。これは、共起関係を導入することで語義の曖昧性解消に有効な組合せが学習できると同時に、ノイズも増えているためであると考えられる。

### 5.2.6 関連研究との比較

表12、および表13は、それぞれ‘line’と‘interest’を用いた関連研究と本手法との比較結果を示す。関連研究における精度は、システムが正しく語義を判定した個数を総正解数で割った値である。

表12において、Mooney (1996)<sup>27)</sup>と Leacockら (1993)<sup>9)</sup>は、ともに多義語を含む文の2文以内にある周辺語を属性として用いている。Mooneyは Naive Bayesian classifierを用い、Leacockらは Neural Networkを用いた結果、それぞれ72%、および76%の精度を得ている。Leacockら (1998)<sup>37)</sup>と Towellら (1998)<sup>20)</sup>は、topical属性とlocal属性を用いている。Leacockらは Naive Bayesian classifierを適用した結果84%の精度を得ている。Towellらは2つの Neural Networkを適用し、分類器が判定した語義の個数が最

表 11 多項式関数の次元と  $F$  値 (SENSEVAL1)  
Table 11  $F$ -measure using the  $d$ -th polynomial functions (SENSEVAL1).

名詞							
語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$
accident	84.6/ <b>85.0</b>	behaviour	95.6/ <b>96.4</b>	bet	71.8/70.0	excess	79.0/78.0
float	93.3/93.3	giant	90.3/89.2	knee	54.9/52.9	onion	98.1/96.7
sack	83.8/82.2	scrap	36.6/35.2	shirt	88.7/86.9		
動詞							
語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$
amaze	98.5/98.2	bet	65.8/62.3	bother	80.4/77.9	bury	70.2/70.1
calculate	77.0/76.4	consume	54.3/52.1	derive	40.7/38.4	float	63.1/61.7
invade	50.2/48.3	promise	75.8/75.3	sack	83.7/ <b>84.2</b>	scrap	94.0/91.9
seize	59.6/59.3						
形容詞							
語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$	語	$d = 1/d = 2$
brilliant	83.7/80.3	floating	59.3/57.4	generous	34.8/33.0	giant	72.2/71.1
modest	82.2/81.2	slight	72.5/71.1	wooden	94.3/ <b>94.8</b>		

表 12 関連研究との比較 'line'

Table 12 Comparison to previous results for 'line'.

WSD 手法	精度
Mooney (1996)	72%
Leacock, Towell & Voorhees (1993)	76%
Leacock, Chodorow & Miller (1998)	84%
Towell & Voorhees (1998)	87%
Pedersen (2000)	88%
本手法 ( $d = 1$ の $F$ 値)	90.0%

表 13 関連研究との比較 'interest'

Table 13 Comparison to previous results for 'interest'.

WSD 手法	精度
Bruce & Wiebe (1994)	78%
Pedersen & Bruce (1997)	78%
Ng & Lee (1996)	87%
Pedersen (2000)	89%
本手法 ( $d = 1$ の $F$ 値)	90.1%

も多い語義に割り当てる voting 手法を用いて解消した結果, 87%の精度を得ている. Pedersen (2000)<sup>28)</sup> は, Naive Bayesian classifier を複数用いることで多義語を解消する手法を提案している. 属性として多義語と周辺語との位置関係に注目し, 多義語からの位置を変化させることで総計 81 個の分類器を作成した. テスト文中の多義語は voting 手法を用い, 88%の精度を得ている.

表 13 は 'interest' を用いた実験結果を示す. Bruce ら (1994)<sup>19)</sup> は, 多義解消に必要な属性は互いに独立ではなく相互に関係していること, しかし, すべての属性どうしの相関関係を考慮にいれた学習を行おうとすると組合せの爆発が起きてしまうことから, 解消に最も有効な属性だけを用い, それらを条件付き確率を用いて表現する Decomposable Model と呼ばれるモデルを提案している. 属性として, 語の形態情

報, 共起情報, および品詞情報を用いた結果, 78%の精度を得ている. Pedersen ら (1997)<sup>38)</sup> は, Bruce らが使用した属性と同じ属性を用い決定木学習を用いた結果, 78%の精度を得ている. Ng ら (1996)<sup>16)</sup> は属性として品詞情報のほか, 派生情報, 多義語の周辺語の情報, 共起関係, および動詞と目的語の構文関係の情報を用いて多義語の解消を行っている. 学習アルゴリズムとして Nearest-Neighbor classifier を適用した結果, 87%の精度を得ている. Pedersen (2000)<sup>28)</sup> は Naive Bayesian classifiers を複数使用し, voting により 89%の精度を得ている.

表 12 と表 13 において, 本手法は 'line', および 'interest' でそれぞれ 90.0%, 90.1%の  $F$  値であったことから, Pedersen の精度 (88%, 89%) と比較可能であるといえる.

SENSEVAL1 を用いた関連研究としては, Kilgarriff らにより SENSEVAL1 コンテストにおける結果が報告されている<sup>34)</sup>. 教師付き学習を用いた手法において最も高い  $F$  値は, 名詞, 動詞, 形容詞でそれぞれ 80%, 70%, および 70~80%であったことが報告されている. 個々のデータに関する詳細な結果が示されていないため正確な比較は不可能であるが, 本手法における名詞, 動詞および形容詞の  $F$  値はそれぞれ 79.5%, 68.3%, および 73.3%であることから, 本手法は既存の手法と同様の精度が得られているといえる.

## 6. おわりに

本稿では, 機械学習 SVM を用いて解消に最適な属性を語義ごとに学習し, これを用いてテスト文中の語義の曖昧性を解消する手法を提案した. 'line' と 'interest', および SENSEVAL1 を用いた実験の結果, 以下の 5 点が明らかになった.

- (1) 本研究で定義した 8 つの属性は, 名詞, 動詞, 形容詞のうち, 名詞の多義解消においてより効果的であった.
- (2) 1 語義あたり 350 文以上の訓練データが存在する語義について, 本手法は高い精度でその語義の曖昧性を解消できる.
- (3) 8 つの属性を用いた属性選択は, 名詞, 動詞, 形容詞のうち, 狭いウィンドウサイズで語義が決まる動詞と形容詞においてより効果的であった.
- (4) 大量の属性を用いると過学習が起こりやすくなるが, 実験結果から属性数の増加にともない必ずしも精度が低下していないことより SVM が高次元の入力属性に対しても高い汎化能力を持つことが分かる.
- (5) ‘line’, ‘interest’, および SENSEVAL1 において多項式の次元が 1 次と 2 次の場合の  $F$  値に差がなかったことから, 多項式の次元数は今回使用したデータにおいて精度に影響を与えていない.

今後の課題として以下の 4 点があげられる.

- (1) SVM は,  $n$  を訓練データ数とすると 1 回の学習に  $O(n^2)$  から  $O(n^3)$  の計算量を要する<sup>33)</sup>. さらに本手法では, 属性の組合せの各々に対して *pairwise* 法を用いて学習しているため, 1 つの多義語の学習に要する時間は, (1 回の学習時間) × (語義の任意の 2 つの組合せ) × (属性の組合せ数) × ( $x$  回の交差検定) となる. 実際に訓練データ数の最も多い多義語である ‘line’ において 1 回の学習に要した時間 (語義 ‘line4’ か ‘line5’ であるかを学習する時間) は, 訓練データ 2,386 文に対して, PentiumIII, 1 GB マシンで 9.42 秒であり, ‘line’ の学習に要した時間は, 54.13 時間であった. 今後は計算量の問題に対処するため, 属性の組合せ手法に対して Wrapper model<sup>39)</sup> や学習のモデル予測<sup>40)</sup> を用いて最適化を行う必要がある.
- (2) 名詞 ‘interest’ が ‘causing attention to be given to’ の意味を持つ場合の  $F$  値は 0% であり, また形容詞 ‘generous’ の場合も極端に精度が低く, 34.8% であった. 原因として, 定義した 8 つの属性の中に解消に有効な属性が含まれていなかったことがあげられる. 今後はこれらの多義語に対しても対処可能な属性を追加する必要がある.
- (3) 実験結果から, 1 語義あたりの訓練データ数が 350 文を超える語義の  $F$  はいずれも 80% を超

- えているのに対し, 300 文以下の語義の  $F$  値の平均は 53.9% であった. 今後は訓練データとテストデータに対して統計手法を用い単語の類似クラスを求めた結果を用いたり, WordNet の同義語クラスである Synset の情報を利用することで訓練データ数が少ない語義に対しても高精度で解消できるよう検討する必要がある.
- (4) 属性選択方法として, Towell らや Pedersen が用いた voting 手法<sup>20),28)</sup> や Wrapper Model<sup>39)</sup> などが提案されている. また, 機械学習も様々な方法が多義解消に適用されている<sup>9),17),27)</sup>. 今後は, 語義ごとの属性選択と SVM との組合せの有効性について検証するため, 語義ごとの属性選択を含むいくつかの属性選択手法と SVM を含むいくつかの機械学習とをそれぞれ組み合わせさせて実験を行い, 定量的に調査することで明らかにする必要がある.

謝辞 本稿に対し, 有益なコメントをいただいた査読者の方々に感謝します. 本研究は, 文部省科学研究費補助金 (財) 人工知能研究振興財団, および通信・放送機構の援助を受けています. ここにそれらを記し謝意に代えさせていただきます.

## 参考文献

- 1) Ide, N. and Véronis, J.: Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art, *Computational Linguistics*, Vol.24, No.1, pp.1–40 (1998).
- 2) Brown, P.: Word-Sense Disambiguation using Statistical Methods, *Proc. 29th Annual Meeting of ACL*, pp.264–304 (1991).
- 3) Dagan, I.: Two Languages are More Informative than One, *Proc. 29th Annual Meeting of ACL*, pp.130–137 (1991).
- 4) Yarowsky, D.: Word-Sense Disambiguation using Statistical Models of Roget’s Categories trained on Large Corpora, *Proc. COLING92*, pp.454–460 (1992).
- 5) Schütze, H.: Dimensions of meaning, *Proc. Supercomputing’92*, pp.787–796 (1992).
- 6) Gale, W., Church, K. and Yarowsky, D.: A method for disambiguating word senses in a large corpus, *Computers and the Humanities*, Vol.26, pp.415–439 (1992).
- 7) Zernik, U.: Train1 vs. Train2: Tagging word senses in a corpus, *Proc. Intelligent Text and Image Handling, RIAO’91*, pp.567–585 (1991).
- 8) Hearst, M.: Noun homograph disambiguation using local context in large corpora, *Proc. 7th Annual Conference of the Centre for the New*

- OED and Text Research: Using Corpora*, pp.1–22 (1991).
- 9) Leacock, C., Towell, G. and Voorhees, E.: Corpus-based Statistical Sense Resolution, *Proc. ARPA Workshop on Human Language Technology* (1993).
  - 10) Yarowsky, D.: Unsupervised Word Sense Disambiguation rivaling Supervised Methods, *Proc. 33rd Annual Meeting of ACL*, pp.189–196 (1995).
  - 11) Fujii, A., Inui, K., Tokunaga, T. and Tanaka, H.: Selective Sampling for Example-Based Word Sense Disambiguation, *Computational Linguistics*, Vol.24, No.4, pp.573–597 (1998).
  - 12) Kilgarriff, A. and Palmer, M.: Introduction, Special Issue on Senseval Evaluating Word Sense Disambiguation Programs, *Computers and the Humanities*, Vol.34, No.1–2, pp.1–13 (2000).
  - 13) RWC: *RWC Text Database* (Japanese), Real World Computing (1995).
  - 14) Black, E.: An Experiment in Computational Discrimination of English Word Senses, *IBM Journal of Research and Development*, Vol.32, No.2, pp.185–194 (1988).
  - 15) Miller, G., Martin, C., Shari, L., Claudia, L. and Thomas, R.: Using a Semantic Concordance for Sense Identification, *Proc. ARPA Workshop on Human Language Technology*, pp.240–243 (1994).
  - 16) Ng, H. and Lee, H.: Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach, *Proc. 34th Annual Meeting of ACL*, pp.40–47 (1996).
  - 17) Stevenson, M. and Wilks, Y.: Combining Weak Knowledge Sources for Sense Disambiguation, *16th Int. Conf. Artificial Intelligence (IJCAI)*, pp.884–889 (1999).
  - 18) Yarowsky, D.: One Sense per Collocation, *Proc. Speech and Natural Language ARPA Workshop*, pp.454–460 (1993).
  - 19) Bruce, R. and Wiebe, J.: Word-Sense Disambiguation using Decomposable Models, *Proc. 32nd Annual Meeting of ACL* (1994).
  - 20) Towell, G. and Voorhees, E.: Disambiguating Highly Ambiguous Words, *Computational Linguistics*, Vol.24, No.1, pp.125–146 (1998).
  - 21) Joachims, T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features, *Proc. Conference on Machine Learning*, pp.96–103 (1998).
  - 22) Yang, Y. and Liu, X.: A re-examination of text categorization methods, *Proc. 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.42–49 (1999).
  - 23) 平 博順, 春野雅彦: Support Vector Machine によるテキスト分類における属性選択, *情報処理学会論文誌*, Vol.41, No.4, pp.1113–1123 (2000).
  - 24) 山田寛康, 工藤 拓, 松本裕治: Support Vector Machines を用いた日本語固有表現抽出, *情報処理学会自然言語処理研究会*, 2001-NL-142, pp.121–128 (2001).
  - 25) 中川哲治, 工藤 拓, 松本裕治: Support Vector Machine を用いた未知語の品詞推定, *情報処理学会自然言語処理研究会*, 2001-NL-141, pp.77–82 (2001).
  - 26) Kudoh, T. and Matsumoto, Y.: Use of Support Vector Learning for Chunk Identification, *Computational Natural Language Learning*, pp.142–144 (2000).
  - 27) Mooney, R.: Comparative Experiments on Disambiguation Word Senses: An Illustration of the Role of Bias in Machine Learning, *Proc. Conference on Empirical Methods in Natural Language Processing*, pp.82–91 (1996).
  - 28) Pedersen, T.: A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation, *Proc. ANLP-NAACL2000*, pp.63–69 (2000).
  - 29) Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer, New York (1995).
  - 30) Burges, C.: *A Tutorial on Support Vector Machines for Pattern Recognition*, Kluwer Academic Publisher (1998).
  - 31) Weston, J. and Watkins, C.: Support Vector Machines for Multi-Class Pattern Recognition, *Proc. 7th European Symposium on Artificial Neural Networks* (1999).
  - 32) Weston, J. and Watkins, C.: Multi-class Support Vector Machines, Technical Report CSD-TR-98-04 (1998).
  - 33) Joachims, T.: Making Large-Scale SVM Learning Practical, *Advances in Kernel Methods — Support Vector Learning*, Schölkopf, B., Burges, C.J.C. and Smola, A.J. (Eds.) (1998).
  - 34) Kilgarriff, A. and Rosenzweig, J.: English SENSEVAL: Report and Results, *Proc. LREC* (2000).
  - 35) Brill, E.: A Simple Rule-based Part of Speech Tagger, *Proc. 3rd Conference on Applied Natural Language Processing*, pp.152–155 (1992).
  - 36) Ramshaw, L. and Marcus, M.: Text Chunking Using Transformation-Based Learning, *Proc. 3rd ACL Workshop on Very Large Corpora*, pp.82–94 (1995).
  - 37) Leacock, C., Chodorow, M. and Miller, G.: Using corpus statistics and WordNet relations

- for sense identification, *Computational Linguistics*, Vol.24, No.1, pp.147–165 (1998).
- 38) Pedersen, T. and Bruce, R.: Distinguishing Word Senses in Untagged Text, *Proc. 2nd Conference on Empirical Methods in Natural Language Processing*, pp.197–207 (1997).
- 39) Kohavi, R. and Sommerfield, D.: Feature subset selection using the wrapper method: Overfitting and dynamic search space topology, *Proc. 1st International Conference on Knowledge Discovery and Data Mining*, pp.192–197 (1995).
- 40) Joachims, T.: Estimating the Generalization Performance of an SVM Efficiently, *Proc. 17th International Conference on Machine Learning*, pp.431–438 (2000).

(平成 12 年 12 月 12 日受付)

(平成 13 年 10 月 16 日採録)



福本 文代 (正会員)

1986 年学習院大学理学部数学科卒業。同年沖電気工業(株)入社。総合システム研究所勤務。1988 年より 1992 年まで(財)新世代コンピュータ技術開発機構へ出向。1993 年マンチェスター工科大学計算言語学部修士課程修了。同大学客員研究員を経て 1994 年より山梨大学工学部助手, 1999 年同学部助教授, 現在に至る。自然言語処理の研究に従事。理学博士。ACM, ACL, 言語処理学会各会員。