

# 決定木学習アルゴリズムを用いた多義語の 訳語選択手法の有効性の評価

水野秀紀      荒木健治      枋内香次

北海道大学大学院工学研究科

〒060-8628 札幌市北区北13条西8丁目

E-mail: {mizuno, araki, tochinai}@media.eng.hokudai.ac.jp

あらまし 我々は、英語から日本語への機械翻訳における多義語の訳語選択に関して、決定木学習アルゴリズムを用いた手法を提案した。本手法は、訳語選択対象語の訳語とその周辺の語の情報の組を訓練事例として決定木学習アルゴリズムを用いて決定木を作成し、作成された決定木を用いて入力文の各単語を分類し、その分類結果から訳語選択を行なう手法である。また、本手法では、従来の事例データに基づく手法で用いられる格情報を用いていないため、構文解析ツールの誤りを考慮する必要がないという利点がある。本稿では、本手法における最適な属性の組合せの評価、そして、本手法と決定木を用いずに訳語選択を行なった場合との比較について述べる。

キーワード 機械翻訳, 多義性解消, 決定木学習アルゴリズム, 事例, シソーラス

## Evaluation of Word Sense Disambiguation Method Using Decision Tree Learning Algorithm

Hideki Mizuno, Kenji Araki, Koji Tochinai

Graduate School of Engineering, Hokkaido University

N13, W8, Kita-ku, Sapporo-shi, 060-8628 Japan

E-mail: {mizuno, araki, tochinai}@media.hokudai.ac.jp

**Abstract** As a new method for word sense disambiguation in machine translation, we have proposed a method using decision tree learning algorithm. In this method, a decision tree is produced by training database. It consists of a set of sense of a multi-meaning word and information of words around the multi-meaning word. Input words are classified using the tree, and an optimum selection of sense of the multi-meaning word is determined by the result of classification. In our method, it is not necessary to consider some errors of parsing tool, because information of case is not used. However, information of case is usually used for the example-based method for word sense disambiguation. In this paper, we evaluate optimum set of the attributes in our method, and compare our method and the method without decision tree learning algorithm.

**key words** machine translation, word sense disambiguation, decision tree learning algorithm, example, thesaurus

# 1 はじめに

近年、様々な機械翻訳システムが研究、開発されている。しかしながら、文に内在する意味的な多義性の解消に関しては、種々の研究 [1][2][3] が行なわれているが、未だ解決されておらず、自然言語処理分野における困難かつ重要な問題となっている。

従来、多義語の訳語選択の手法は、文法的な解析に基づく手法 [1] が主であった。この手法は、文法規則などを基に訳語選択を行なう手法で、文法的に正しい結果が得られるが、多様な言語現象を有限個の文法規則で記述することの困難さや文法規則や辞書の巨大化に伴うそれらの作成や保守・改良の困難さが問題点として挙げられる。

このような問題点を解決するため、最近、事例データに基づく手法 [2][3] が開発されている。この手法は、入力文と事例データとの間で何らかの類似度を計算して、最も類似度の高い事例を用いて訳語選択を行なう手法で、事例データが解析された文の集合であるため、容易に事例の追加や削除ができ、複雑な文にも対応できる。

しかしながら、事例データに基づく手法には、大量の事例を必要とすることや、類似度の定義法に多くのバリエーションが考えられ、最適な手法が存在しないという問題点がある。また、事例データに基づく手法では格情報を用いているものが多いが、格情報を用いるためには構文解析ツールを用いる必要があり、その誤りを考慮する必要がある。また、訳語選択対象語の周辺の語の多義性が解消されていなければ訳語選択が行なえないという問題点も挙げられる。

そこで我々は、人間が訳語選択対象語の周辺の語に注目して経験的則則から訳語選択を行なっているという仮定に基づき、その状況を計算機上に実現するために、決定木学習アルゴリズム C4.5[4] を用いた手法を提案した [5]。本手法では、訳語選択対象語の訳語をクラス、その周辺の語の情報を属性として決定木学習アルゴリズム C4.5[4] を用いて決定木を作成する。次に、作成された決定木を用いて入力文の訳語選択対象語の周辺の語の分類を行ない、その分類結果を基に訳語選択を行なう手法である。本手法は、事例データに基づく手法であるが、訳語選択対象語の周辺の語の情報に従来の手法に見られるような格情報を用いていないので構文解析ツールの誤りを考慮する必要がないという利点を有する。

本稿では、本手法における最適な属性の組合せの評価、そして、本手法と決定木を用いずに訳語選択を行なった場合との比較について述べる。

# 2 処理過程

## 2.1 概要

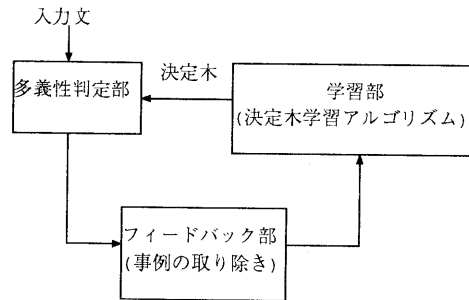


図 1: 処理過程

本手法に基づくシステムの処理過程を図 1 に示す。以下、処理の概要を述べる。まず、訳語選択対象語を 1 つ定め、その語を含む英文をシステムに入力する。入力された英文の各単語に対して、学習部で作成された決定木を基に多義性判定部において分類を行なう。そして、その分類結果を基に判定式を用いて最適解の選択を行なう。その後、システムの出力した最適解に対しユーザーが正解を与え、入力された英文の各単語を訓練事例に加え、加えられた訓練事例がフィードバック部へ渡される。次に、フィードバック部において事例の取り除き処理を行ない、取り除き処理を行なった訓練事例が学習部へと渡される。そして、学習部においてフィードバック部から渡された訓練事例を用いて決定木の作成を行ない、作成された決定木が再び多義性判定部へ渡される。

## 2.2 事例の構成

本手法では、図 2 のような、訳語選択対象語  $m_S$  を含む文  $S$  において、 $S$  を構成する 1 単語  $s_i$  を 1 つの事例とし、各事例がその事例の情報を表す属性とその属性値、その事例が属する文での訳語選択対象語の意味を表すクラスを持つ。但し、訳語選択対象語  $m_S$  そのものは事例に含まれない。また、事例となる単語の多義性の解消は行なわない。

$$S: s_1 s_2 \dots s_i \dots s_{j-1} m_S s_j \dots s_n$$

$\leftarrow D_{s_i} \rightarrow$

図 2: 文

本手法で用いた属性の一覧を表 1 に示す。表 1 におい

表 1: 属性の一覧

属性番号	内容	属性番号	内容
1	訓練事例における頻度	20	contact (動)
2	訳語選択対象語からの文上の距離	21	creation (動)
3	entity (名)	22	emotion (動)
4	psychological feature (名)	23	motion (動)
5	abstraction (名)	24	perception (動)
6	location (名)	25	possession (動)
7	shape form (名)	26	social (動)
8	state (名)	27	stative (動)
9	event (名)	28	weather (動)
10	act, human action (名)	29	all (形)
11	group, grouping (名)	30	pert (形)
12	possession (名)	31	ppl (形)
13	phenomenon (名)	32	all (副)
14	body (動)	33	代名詞
15	change (動)	34	前置詞
16	cognition (動)	35	助動詞
17	communication (動)	36	接続詞
18	competition (動)	37	字面
19	consumption (動)		

て、属性番号 2 の属性値には、式 (1) により求めた  $L_{s_i}$  を離散化して用いる。式 (1) において、 $D_{s_i}$  は、文  $S$  中の語  $s_i$  が訳語選択対象語  $m_S$  から何語離れているかを表し、 $A_S$  は、文  $S$  中の訳語選択対象語  $m_S$  より前にある語でシステムに入力される語数を表し、 $B_S$  は、文  $S$  中の訳語選択対象語  $m_S$  より後ろにある語でシステムに入力される語数を表す。

$$L_{s_i} = \begin{cases} -\frac{D_{s_i}}{A_S} & (s_i \text{ が } m_S \text{ の前にある場合}) \\ \frac{D_{s_i}}{B_S} & (s_i \text{ が } m_S \text{ の後ろにある場合}) \end{cases} \quad (1)$$

また、属性番号 3 から 13 は、それぞれシソーラス WordNet[6] における最上位概念を表し、属性番号 14 から 32 は、それぞれシソーラス WordNet[6] の辞書ファイルに登録されている分類名を表し、属性番号 33 から 36 は、それぞれシソーラス WordNet[6] には登録されていない品詞を表す。そして、属性番号 3 から 32 の属性値には、それぞれの属性を取り得る確率を離散化して用い、属性番号 33 から 36 の属性値には、それぞれの属性を取り得るか取り得ないかの有無を用いる。

## 2.3 学習部

学習部では、決定木学習アルゴリズム C4.5[4] を用いて訓練事例集合から決定木の作成を行なう。決定木学習アルゴリズム C4.5[4] では、以下の 2 段階の手続きによって決定木の作成を行なう。

1. 属性、属性値、クラスから成る事例集合が与えられると、情報理論的なヒューリスティック関数を最大にする属性を選んで事例集合を再帰的に分割し、クラスを完全に弁別できる決定木を作成する。決定木の各節点はある属性による事例の分割を表し、その属性値に応じて分岐する。各節点は後の処理のために各々に割り当てられた事例のクラス分布を保持している。分割は、決定木の葉に出来る限り 1 つのクラスの事例が集中するように行なわれる。
2. 1 で作成された決定木は、訓練事例に過適応しており、一般的な分類能力が必ずしも高いとは限らない。そこで予測誤り率に基づく決定木の枝刈りを行なう。

## 2.4 多義性判定部

入力文中でシステムに入力される単語の集合を  $X$  としたとき、多義性判定部では、学習部で作成された決定木を基に  $X$  に含まれる単語  $x_i$  をそれぞれ分類して、それぞれのクラスに属する確率を求める。その際、単語  $x_i$  と同じ字面の単語が訓練事例に存在しない場合は、シソーラス WordNet[6] から単語  $x_i$  の訓練事例に存在する同義語・上位語・下位語を抽出し、抽出した語で分類を行なう。但し、単語  $x_i$  と全ての属性に関してその属性値の等しい事例が後述するフィードバック部で既に取り除かれている場合は、単語  $x_i$  は分類を行わない。

そして、式(2)によりクラス  $C_j$  のスコアを求め、スコアの値が最も大きいクラス  $C_j$  を入力文の最適解として選択する。式(2)において、 $P_{x_i}(C_j)$  は、入力中の語  $x_i$  が決定木によりクラス  $C_j$  に分類される確率を表し、 $w(x_i, y_i)$  は、入力中の語  $x_i$  と実際に分類を行なった語  $y_i$  との間のシソーラス WordNet[6] 上の距離に応じた重み係数を表す。

$$Score(C_j) = \sum_{x_i \in X} w(x_i, y_i) P_{x_i}(C_j) \quad (2)$$

なお、シソーラス上の距離については、シソーラス WordNet[6] の概念階層を基に次のように定義する [7]。

[定義1] 任意の語とそれに隣接する上位語（下位語）との間の距離は全て等しい。

[定義2] 任意の語とその同義語との間の距離は全て等しい。

[定義3] 任意の語とそれ以外の任意の語と間の距離は共通する上位語までのそれぞれの距離の和とする。

この定義に従って、次のような例を考えてみる。シソーラス WordNet[6] から図3のような概念階層の一部が取り出されると、device の1つ上の上位語 instrumentality であり、device の1つ下の下位語が machine と fan であるから、 $d(x_1, x_2)$  を語  $x_1$  と語  $x_2$  とのシソーラス上の距離とすると、上述の [定義1] と [定義3] より、

$$\begin{aligned} d(\text{device}, \text{instrumentality}) &= d(\text{machine}, \text{device}) \\ &= d(\text{fan}, \text{device}) \end{aligned} \quad (3)$$

$$\begin{aligned} d(\text{machine}, \text{fan}) &= d(\text{machine}, \text{device}) + d(\text{fan}, \text{device}) \\ &= 2d(\text{machine}, \text{device}) \end{aligned} \quad (4)$$

となる。

また、system と unit は同義語であり、instrumentality と instrumentation も同義語であるから、上述の [定義2] より、

$$\begin{aligned} d(\text{system}, \text{unit}) &= d(\text{instrumentality}, \text{instrumentation}) \end{aligned} \quad (5)$$

となる。

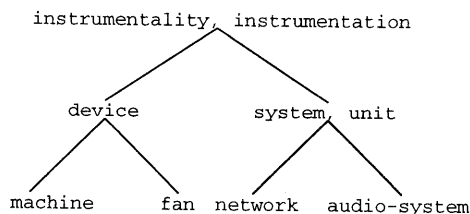


図3: シソーラス

## 2.5 フィードバック部

フィードバック部では、事例を適応的に構築するために事例の取り除き処理を行なう。取り除きの方法は、クラスの異なる同じ単語があった場合にその単語同士を表1の属性番号2の属性値の差を計算し、その逆数をその単語の取り除き指数に加えていく。そして、取り除き指数がある閾値を越えたらその単語は事例から取り除かれる。このことにより、全ての属性に関してその属性値が等しいのにクラスのみが異なっている事例が取り除かれ、決定木の精度が向上すると考えられる。

## 3 評価実験

### 3.1 実験方法

2章で述べたシステムを用いて以下の点について評価実験を行なった。

1. 本手法における最適な属性の組合せの抽出とその評価。
2. 本手法と決定木を用いずに字面情報のみを用いて訳語選択を行なった場合との比較。

本実験は、訳語選択対象語を take とし、事例の初期状態を空にしてからプログラム言語のマニュアル

表 2: 実験結果

実験条件	属性値のグループ化を行なわなかった場合				属性値のグループ化を行なった場合			
	正解 [文]	誤り [文]	不能 [文]	正解率 [%]	正解 [文]	誤り [文]	不能 [文]	正解率 [%]
(1)	177	98	1	64.4	193	83	0	69.9
(2)	192	84	0	69.6	191	85	0	69.2
(3)	174	101	1	63.3	204	72	0	73.9
(4)	179	96	1	65.1	190	86	0	68.8
(5)	185	91	0	67.2	190	86	0	68.8
(6)	192	83	1	69.8	198	77	1	72.0
(7)	178	97	1	64.7	204	72	0	73.9
(8)	188	88	0	68.1	191	85	0	69.2
(9)	177	98	1	64.4	194	81	1	70.5
(10)	177	99	0	64.1	177	99	0	64.1
(11)	174	101	1	63.3	203	73	0	73.6
(12)	177	98	1	64.4	194	81	1	70.5
(13)	171	105	0	62.0	176	100	0	63.8
(14)	162	113	1	58.9	167	109	0	60.5
(15)	174	101	1	63.3	204	72	0	73.9

[8][9][10][11] 内の take を含む英文 280 文を用いて行なった。

また、本実験で用いる決定木は、属性値のグループ化を行なったものを行なわなかったものの 2 種類を用いた。

また、入力文の全単語を入力するのではなく、あらかじめ行なった予備実験より take の前後 3 語を入力した。

また、冠詞は、take の多義性を解消する際に関与しないと考えられるので、あらかじめ除外をしておいた。

また、式 (2) の重み係数に関しては、2.4 節で述べたシソーラス上の距離の定義に従って、任意の語とそれに隣接する上位語（下位語）との間の距離に 1 という値を与え、任意の語とその同義語との間の距離に 0.5 という値を与え、式 (6) で求めた  $w(x_i, y_i)$  を重み係数として用いた。式 (6) において、 $x_i$  を入力文中の語、 $y_i$  を実際に分類を行なった語とした時、 $d(x_i, y_i)$  は  $x_i$  と  $y_i$  とのシソーラス上の距離を表す。

$$w(x_i, y_i) = 10 - d(x_i, y_i) \quad (6)$$

なお、あらかじめ行なった予備実験より、take の前後 2 語に関しては式 (6) で求めた値を 3 倍して重み係数として用い、その他の語に関しては式 (6) で求めた値をそのまま重み係数として用いた。

表 3: 実験条件

実験条件	内容
(1)	全ての属性番号を使用
(2)	属性番号 1 から 36 を使用
(3)	属性番号 1 と 3 から 37 を使用
(4)	属性番号 2 から 37 を使用
(5)	属性番号 1 と 3 から 36 を使用
(6)	属性番号 2 から 36 を使用
(7)	属性番号 3 から 37 を使用
(8)	属性番号 3 から 36 を使用
(9)	属性番号 1, 2, 37 を使用
(10)	属性番号 1, 2 を使用
(11)	属性番号 1, 37 を使用
(12)	属性番号 2, 37 を使用
(13)	属性番号 1 のみを使用
(14)	属性番号 2 のみを使用
(15)	属性番号 37 のみを使用

### 3.2 実験結果

実験結果を表2に、表2の実験条件を表3に示す。表2において、正解はシステムの出力した最適解と正解となるべきクラスとが一致したものを表し、誤りはシステムの出力した最適解に正解となるべきクラスが含まれていないものを表す。また、不能はシステムの出力した最適解に正解となるべきクラスが含まれているものの、その最適解に複数の解が出力されてしまったものを表す。正解率は式(7)で計算された値を表す。

$$\text{正解率} [\%] = \frac{\text{正解の数}}{\text{正解の数} + \text{誤りの数}} \times 100 \quad (7)$$

いずれの実験条件においても、訓練事例に正解となるべきクラスが登録されていなかったものが4文あった。

## 4 考察

### 4.1 最適な属性の組合せの検討

表2より、最も正解率の高かった実験条件は、実験条件(3)、(7)、(15)の属性値のグループ化を行なったものとなった。また、属性に属性番号37の字面を含み、属性値のグループ化を行なったものは比較的良好な結果が得られた。このことから、本手法では字面を属性として含み、属性値のグループ化を行なうと良い結果が得られることが分かった。

また、逆に、属性として字面が含まれていても属性値のグループ化を行なわなかった場合は、いずれも良い結果が得られていない。これは、字面が取り得る属性値が他の属性に比べ非常に多く、字面による事例の分割を表す節点において、多数の分岐ができてしまい、精度の良い決定木が作成されないためであると考えられる。

### 4.2 決定木を用いなかった場合との比較

本手法と決定木を用いずに、字面情報のみを用いて訳語選択を行なった場合との比較を行なうために、以下のような処理過程で実験を行なった。実験方法は、3.1節で述べた方法に従った。

- 学習部では、決定木を作成しない。
- 多義性判定部では、決定木による分類の代わりに、式(8)により式(2)の  $P_{x_i}(C_j)$  を求め、式(2)を用いて最適解を選択する。式(8)において、 $V_{x_i}(C_j)$  は、入力される単語  $x_i$  と同じ字面の訓練事例の単

語の中でクラスが  $C_j$  となっているものの個数を表し、 $T_{x_i}$  は、入力される単語  $x_i$  と同じ字面の単語の訓練事例での個数を表す。入力される単語  $x_i$  と同じ字面の単語が訓練事例に存在しない場合は、2.4節で述べたように、シソーラス WordNet[6] から単語  $x_i$  の訓練事例に存在する同義語・上位語・下位語を抽出し、抽出した語で分類を行なう。また、式(2)の重み係数  $w(x_i, y_i)$  は、2.4節で述べた方法と同様にして求める。

$$P_{x_i}(C_j) = \frac{V_{x_i}(C_j)}{T_{x_i}} \quad (8)$$

- フィードバック部では、2.5節で述べた方法と同様な処理を行なう。

実験の結果を表4に示す。また、決定木を用いることによる改善率の推移を図4に、決定木の使用の有無による正解と誤りの内訳を表5に、具体例を以下に示し、その出力結果を表6、表7、表8、表9に示す。

[例1] 決定木の使用で誤りから正解となった例

This statement takes a list of values.

[例2] 決定木の使用で正解から誤りとなった例

The types, if otherwise identical, are also taken to agree.

なお、決定木を用いた方は、表2において、最も正解率の高かった実験条件(15)の属性値のグループ化を行なった場合である。

表4: 決定木を用いなかった場合の実験結果

正解 [文]	誤り [文]	不能 [文]	正解率 [%]
198	78	0	71.7

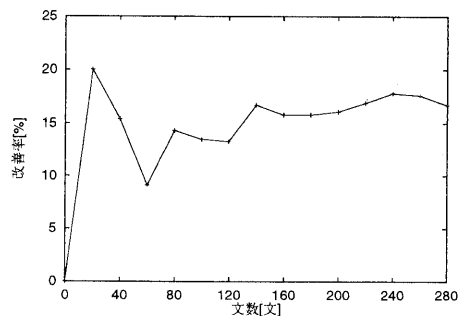


図4: 改善率の推移

表2において最も正解率の高かった実験条件(15)の属性値のグループ化を行なった場合の正解率と表4の正解率を比較すると、わずかではあるが、前者の正解率の方が高いのが分かる。これは、例1の表6と表7のstatementの分類結果を見ると、statementは訓練事例にクラス3のみ存在するので、表7では、クラス3の $P_{x_i}(C_j)$ の値が1.00となっているが、表6では、 $P_{x_i}(C_j)$ の値がクラス3で0.96、クラス1で0.04となっていることから、決定木を用いることにより、ある程度の曖昧性を考慮した分類を行なうことが可能になり、正解率が上昇したと考えられる。また、図4より、280文での改善率は、16.7%となった。

表5: 決定木の使用の有無による正解と誤りの内訳

決定木の使用		文数[文]
有	無	
正解	誤り	13
誤り	正解	7

表6: 例1で決定木を用いた場合の出力結果

入力語	重み係数	クラス	$P_{x_i}(C_j)$
this	0		
statement	30.0	1	0.04
		3	0.96
list	30.0	1	0.86
		2	0.02
		3	0.06
		4	0.06
of	0		
value	10.0	1	0.86
		2	0.02
		3	0.06
		4	0.06
出力クラス			1
正解クラス			1

表7: 例1で決定木を用いなかった場合の出力結果

入力語	重み係数	クラス	$P_{x_i}(C_j)$
this	0		
statement	30.0	3	1.00
list	30.0	1	0.80
		3	0.20
of	0		
value	10.0	1	1.00
出力クラス			3
正解クラス			1

表8: 例2で決定木を用いた場合の出力結果

入力語	重み係数	クラス	$P_{x_i}(C_j)$
identical	0		
be	0		
also	0		
to	30.0	1	0.03
		3	0.03
		4	0.93
agree	30.0	1	0.17
		3	0.81
		4	0.02
出力クラス			4
正解クラス			3

表9: 例2で決定木を用いなかった場合の出力結果

入力語	重み係数	クラス	$P_{x_i}(C_j)$
identical	0		
be	0		
also	0		
to	30.0	1	0.29
		3	0.46
		4	0.25
agree	30.0	3	1.00
出力クラス			3
正解クラス			3

一方、例2の表8と表9のtoの分類結果を見ると、訓練事例にはクラス3のtoが最も多く存在するため、表9では、クラス3の $P_{x_i}(C_j)$ の値が最も大きいが、表8では、クラス4の $P_{x_i}(C_j)$ の値が最も大きくなり、決定木を用いることにより正しい分類が行なわれず、誤りとなっている。

### 4.3 誤りの原因の検討

表10: 誤りの原因

原因	文数[文]	割合[%]
スパース性	55	76.4
分類結果が誤り	10	13.9
重み係数	7	9.7
合計	72	100.0

表2において、最も正解率の高かった実験条件(15)の属性値のグループ化を行なった場合の中で誤りとなってしまったものの原因を表10に示す。

表10より、スパース性が原因で誤りとなったものが最も多いことが分かる。これは、入力文のクラスと、入

力される単語と同じ字面の訓練事例の単語のクラスとが一致していないのが原因であり、この問題を解決するためには、訓練事例の数を増やす必要があると考えられる。

#### 4.4 他手法との比較

本手法と従来の手法とを比較した場合、決定木学習アルゴリズム C4.5 を用いた手法として田中の手法 [12] があるが、この手法では、動詞 take に関する 250 個の未学習事例の誤訳率が 26.0% という結果が示されている。本手法と同じデータを用いていないため、明確な比較はできないが、田中の手法 [12] では、人手により格要素ラベルの付与、主辞の認定、主辞の日本語訳語の付与を行なっていて、これらを人手により行なわなかった場合の影響が考慮されていない点や、訳語選択対象語の周辺の語の多義性の解消が必要であるという問題点が挙げられる。また、事例データに基づく単語多義性解消手法に関して、藤井ら [3] が比較を行なっているが、最も精度の高い手法の場合、日本語コーパス 10,800 文を用いた正解率が 62.6% と示されていて、あまり高い精度が得られていない。

これに対し、我々の手法は、格情報を用いず、また、訳語選択対象語の周辺の語の多義性の解消は行なっていないにもかかわらず、73.9% の正解率が得られている点から、分野がプログラム言語と限定されているが、従来の手法より有望であると考えられる。

## 5 おわりに

本稿では、多義語の訳語選択対象語の周辺の語の情報から決定木学習アルゴリズムを用いて訳語選択を行なう手法の有効性を評価実験の結果より述べた。その結果、決定木学習アルゴリズムを用いることによって、16.7% の改善率が得られた。また、本手法では、字面を属性として含み、決定木の作成の際に、属性値のグループ化を行なうと良い結果が得られることが分かった。また、本手法は、他の従来の手法と比べて、格段に精度が高いわけではない。しかし、本手法は、事例データに基づく手法で多く用いられる格情報を用いず、また、訳語選択対象語の周辺の語の多義性の解消を行なわなくても 73.9% の正解率が得られ、有望な手法であると考えられる。

今後の課題として、今回、最適な属性の組合せの検討の際に、属性番号 3 から 36 を一まとまりにして扱ったので、それらをバラバラにした最適な属性の組合せの検討や、フィードバック部の有効性の検討を行ないたいと考えている。また、4.3 節で最大の誤りの原因と

なったスパース性の問題を解決するために、さらに文の数を増やして実験を行ないたいと考えている。また、格情報を全く用いていないので、話し言葉などの非文の多い分野での評価実験も行ないたいと考えている。

## 謝辞

本研究の一部は文部省科学研究費補助金（課題番号 10680367）により行なわれた。

## 参考文献

- [1] 野村浩郷編, “言語処理と機械翻訳,” 講談社, 1991.
- [2] Naohiko Uramoto, “Example-Based Word-Sense Disambiguation,” *IEICE Trans. Inf. & Syst.*, vol.E77-D, no.2, pp.240-246, Feb. 1994.
- [3] 藤井 敦, 乾健太郎, 徳永健伸, 田中穂積, “単語多義性解消法の比較検討,” *情処学 NL 研報*, NL119-8, pp.45-52, May 1997.
- [4] Quinlan, J. R., “*C4.5: Programs for Machine Learning*,” Morgan Kaufmann, 1993.
- [5] 水野秀紀, 荒木健治, 宮永喜一, 枋内香次, “決定木学習アルゴリズムを用いた多義語の訳語選択手法,” 第 56 回情処学全大, 分冊 2, no.6Q-9, pp.283-284, March 1998.
- [6] Christiane Fellbaum, ed., “*WordNet: An electronic lexical database*,” The MIT Press, 1998.
- [7] 水野秀紀, 荒木健治, 宮永喜一, 枋内香次, “機械翻訳における事例を用いた多義語の訳語選択手法,” *情処学 NL 研報*, NL120-2, pp.7-13, July 1997.
- [8] Brian W. Kernighan, and Dennis M. Ritchie, “*The C programming language. Second Edition*,” Prentice-Hall, 1988.
- [9] Alfred V. Aho, Brian W. Kernighan, and Peter J. Weinberger, “*The AWK programming language*,” Addison-Wesley Pub. Co., 1988.
- [10] Randal L. Schwartz, “*Learning Perl. Second Edition*,” O’Reilly & Associates, 1997.
- [11] Larry Wall, and Randal L. Schwartz, “*Programming perl*,” O’Reilly & Associates, 1992.
- [12] 田中英輝, “木構造の属性を許す決定木学習アルゴリズム,” *情処学論*, vol.38, no.11, pp.2122-2133, Nov. 1997.