

平成27年度茨城大学工学部情報工学科
卒業研究論文

辞書の語義を見出しとした概念辞書を用いた
教師なし語義曖昧性解消

平成28年2月9日提出
茨城大学 工学部情報工学科
12T4069L 佐鳥 恭太郎
指導教員：新納 浩幸 教授

辞書の語義を見出しとした概念辞書を用いた 教師なし語義曖昧性解消

氏名：12T4069L 佐鳥 恭太郎

指導教員：新納 浩幸 教授

論文要旨

本論文では辞書の語義を見出しとした概念辞書を用いた教師なし語義曖昧性解消を行う。

教師なし機械学習手法による語義曖昧性解消は all-words WSD に利用できるため活発に研究されている。ただし、多くの場合、そこでの語義は概念である。それは教師なし機械学習手法では、何らかのヒューリスティックを識別の手がかりとしており、単語間に曖昧性があるために、語義間の距離が必要になる。このような背景から、新納らは辞書の語義を見出しとした概念辞書を構築することで、辞書の語義を付与する all-words WSD を実現している。

ただし、そこには2つの問題がある。1つは語義の識別に用いている手法が概念の uni-gram と bi-gram だけという単純なものであることである。語義の識別の手がかりとなる単語は2単語以上離れた位置に出現することもあり得る。もう1つの問題は上記の uni-gram と bi-gram を評価データから構築していることである。教師なし学習手法では、概念と単語間の共起の強さが必要だが、その強さを測るのに評価データを用いた場合、そのデータに合った共起の強さが得られるために、実際の評価時に有利に働く可能性がある。つまり、公正な評価が行われていない可能性がある。

本論文では、上記2点を改善した教師なし語義曖昧性解消を行う。概念辞書は新納らが作成したものを利用する。ただし、概念と単語間の共起の強さを測るために、前後の単語だけでなく1文内の共起単語を対象にする。また、概念と単語間の共起の強さを測るために用いるコーパスを、評価データとは別種のコーパス（'95年版毎日新聞記事）を用いる。概念と単語間の共起の強さから語義曖昧性解消を行う手法としては Yarowsky の手法を利用する。

実験では構築した概念と単語間の共起の強さを利用して SemEval-2 の日本語辞書タスクのテスト用例の曖昧単語の語義識別を行ったところ、マイクロ平均の正解率として 46.08 % を得た。

本論文では、新納の手法で問題であった語義識別の手掛かりを見逃してしまう手法の問題と、共起の強さを測るために評価データを用いた問題の2つを解決した。しかし、正答率は高いとは言えず、正答率の向上が今後の課題である。

目次

第 1 章	はじめに	1
1.1	概要	1
1.2	本論文の構成	2
第 2 章	語義曖昧性解消	3
2.1	概要	3
2.2	語義曖昧性解消の手法	4
第 3 章	概念辞書を利用した教師なし WSD	5
3.1	概要	5
3.2	概念辞書の作成	5
3.2.1	概念辞書とクラスタリング	5
3.2.2	半教師クラスタリング	6
3.2.3	作成された概念辞書	6
3.3	問題点	7
第 4 章	Yarowsky の手法	8
4.1	概要	8
4.2	共起の強さの計算	9
4.3	語義曖昧解消	10
第 5 章	実験	11
5.1	概要	11
5.2	共起の強さ	12
5.3	語義曖昧性解消	12
5.4	評価データ	13
5.5	結果	13

第 6 章 考察	15
6.1 他手法との比較	15
6.1.1 baseline システム	15
6.2 新納の手法の問題解決	16
第 7 章 結論	17
謝辞	18
参考文献	19

表 目 次

2.1 「ゆく」の語義	3
4.1 共起の強さの合計	10
5.1 曖昧単語	13
5.2 単語の正答率	14
6.1 素性	15

目 次

3.1	概念辞書 (一部)	6
4.1	”ゆく”の持つ語義と語義の属するクラスター	9
4.2	共起の強さ計算	9

第1章 はじめに

1.1 概要

本論文では辞書の語義を見出しとした概念辞書を用いた教師なし語義曖昧性解消を行う。

教師なし機械学習手法による語義曖昧性解消は all-words WSD に利用できるため活発に研究されている。教師なし語義曖昧性解消では、多くの場合概念辞書を用いて語義を概念に変換している。新納らは辞書の語義を見出しとした概念辞書を構築することで、辞書の語義を付与する all-words WSD を実現している [4]。

しかし、問題点が2つある。1つは語義の識別に用いている手法が概念の uni-gram と bi-gram だけという単純なものであることである。語義の識別の手がかりとなる単語は2単語以上離れた位置に出現することもあり得る。もう1つの問題は上記の uni-gram と bi-gram を評価データから構築していることである。教師なし学習手法では、概念と単語間の共起の強さが必要だが、その強さを測るのに評価データを用いた場合、そのデータに合った共起の強さが得られるために、実際の評価時に有利に働く可能性がある。つまり、公正な評価が行われていない可能性がある。

本論文では、上記2点を改善した教師なし語義曖昧性解消を行う。概念辞書は新納らが作成したものを利用する。概念と単語間の共起の強さを測るために、1文内の共起単語を対象とする。共起単語を前後の単語だけでなく1文内の単語に広げることで、新納の手法において問題であった語義識別の手掛かりを見逃すことを解決する。また、概念と単語間の共起の強さを測るために用いるコーパスを、評価データとは別種のコーパス（'95年版毎日新聞記事）を用いる。共起の強さを測るために用いるコーパスと評価データを別のデータにすることで、新納の手法において問題であった、共起の強さが評価データに有利になることを解決する。概念と単語間の共起の強さから語義曖昧性解消を行う手法として Yarowsky の手法 [2] を利用する。

実験では、概念辞書を用いて'95年版毎日新聞から共起の強さを構築した。共起する概念は、SemEval-2 の日本語辞書タスク [1] のテスト用例の曖昧単語 50

単語の語義を対象とした．構築した概念と単語間の共起の強さを利用し曖昧単語の語義識別を行った．よって，新納の手法の問題点2つを解決して，対象の曖昧単語について語義曖昧性解消を行うことができた．

1.2 本論文の構成

2章の語義曖昧性解消 (Word Senese Disambiguation, WSD) から始まり，3章の all-words WSD を行うために新納らが作成した概念辞書の自動作成の手法について，4章の語義曖昧性解消を行う手法の1つである，Yarowsky の手法について概説する．

そして，5章にて新納らの作成した概念辞書を利用して Yarowsky の手法にて新納らの手法の問題点を解決した語義曖昧性解消を行った実験を概説する．6章では，baseline システムとの比較や新納らの手法の問題解決について述べる．7章では結論と今後の課題を述べる．

第2章 語義曖昧性解消

2.1 概要

複数の語義(意味)を持つ単語が文中に存在したときに, その単語が持つべき語義を決定するタスクである. これを語義曖昧性解消 (Word Sense Disambiguation, WSD) といいう.

例として「ゆく」の語義を例として表 2.1 に示す. 文「私は茨城大学へゆく。」が存在するとき, 文中に現れる「ゆく」は表 2.1 の (1) の語義である. このように多義語に対して文脈に合った語義を決定することが語義曖昧性解消である.

表 2.1: 「ゆく」の語義

単語	語義
ゆく	(1) 目的地へ向かって進む (2) 年月が経過する (3) 物事がはかどる

2.2 語義曖昧性解消の手法

語義曖昧性解消は、一般的にデータを読み込み、何かしらの手法によって単語に正しい語義を決定するように学習したモデルを用いて行う。

学習手法は主に教師あり学習，半教師あり学習と教師なし学習が存在する。学習には，データとしてコーパスが与えられる。コーパスとは，文書データを学習しやすいように情報を付加したデータである。情報とは，文の形態素解析や構文解析の結果や多義語に対して語義タグなどである。この中でも，形態素解析と構文解析はソフトウェアが公開されているので，簡単に付加することができる。一般的にコーパスを訓練データとして学習し，評価データのコーパスを用いて学習の優劣を評価する。以下に教師あり学習，半教師あり学習と教師なし学習の特徴 [3] を示す。

- | | |
|---------|--|
| 教師あり学習 | <ul style="list-style-type: none">・ 十分な情報付き訓練データと少量の情報付き評価データを用いて学習する。・ 訓練データから学習した結果を評価データで評価して，学習を最適化するので良い結果が得られやすい。・ しかし，十分な情報付き訓練データを集めるのは高コスト。 |
| 半教師あり学習 | <ul style="list-style-type: none">・ 十分な情報付き訓練データと十分な情報なし評価データを用いて学習する。・ 情報付きと情報なしのデータを混ぜることで，情報付きだけで学習したときよりもよい結果を得ることが目的。 |
| 教師なし学習 | <ul style="list-style-type: none">・ 情報付き訓練データのみで学習する。・ 訓練データに付加する情報が人手を用いない場合，文書データを集めるだけで済むため，コストは無いと見てよい。 |

第3章 概念辞書を利用した教師なし WSD

3.1 概要

all-words WSD を行うためには教師なし学習手法が有望である。この手法での語義は概念であり，概念は概念辞書に記載されている。そこで，新納らは語義を見出しとした概念辞書の自動作成を目的とする論文を発表している。

新納らが提案した手法では，語義タグ付きコーパスから語義を見出しとした概念辞書を作成し，概念 n-gram を利用することで all-words WSD を行っている。

3.2 概念辞書の作成

3.2.1 概念辞書とクラスタリング

ある単語の持つ複数の語義は，それぞれ異なる概念に属していると仮定すると，単語と概念が決定すると対応する語義が決定する。つまり，この仮定は語義をそれぞれ異なるクラスターに分けるクラスタリングと同義である。

新納らの手法では，Word2Vec により 語義付きコーパスを単語クラスタリングしている。Word2Vec とは，単語をベクトル化して表現する手法である。また，オプションを指定することで単語をクラスタリングすることも可能である。


クラスタリングの方法は，語義タグ付きコーパスに存在する語義タグが付いている単語は語義タグを，語義タグが付いていない単語は単語をそれぞれ取り出し，語義と単語の分かち文を作成する。この様に語義を単語とみなすことで，語義のクラスタリングを行うことができる。

3.2.2 半教師クラスタリング

Word2Vec による語義のクラスタリングでは、1つのクラスタに同じ単語の語義同士が属してしまう問題がある。この問題は概念辞書の条件である、「1つの概念に同じ単語の語義は含まれない」に反している。この問題の解決法として、同じクラスタに属する同じ単語の語義を、さらにクラスタリングすることで別のクラスタへ分ける。つまり、同じ単語の語義は同じクラスに属さないという制約をつけた半教師ありクラスタリングをするということである。以上の操作により、新納らによって語義を見出し語とする概念辞書が作成された。

3.2.3 作成された概念辞書

新納らによって作成された語義を見出しとした概念辞書の一部を図 3.1 に示す。画像の見方としては、「行数 | 語義 クラスタ」となっている。語義は、「単語番号-語義番号-細分類...」で表現されている。多義語それぞれに番号がついており、語義と語義の細分類が「-(ハイフン)」で区切られている。この概念辞書は一行に語義と属するクラスタが記されている。半教師クラスタリングにより、クラスタ「4」が「4.0」と「4.1」に分かれていることが見て取れる。これにより、単語「1460」の語義「1460-0-0-1」と「1460-0-0-2」が同じクラスタに属さないようにしている。このようなクラスタが 26 クラスタ存在しており、概念辞書のクラスタ数の合計は 526 クラスタとなっている。



41	6599-0-0-3	3↓
42	1460-0-0-1	4.1↓
43	1460-0-0-2	4.0↓
44	19086-0-0-2	4.0↓
45	34498-0-0-2	4.0↓
46	47954-0-0-1	4.0↓
47	4810-0-0-1	4.0↓
48	33144-0-0-2	5↓
49	21387-0-0-0	6↓

図 3.1: 概念辞書 (一部)

3.3 問題点

新納らの手法には、2つの問題点がある。1つは、語義の識別に、単純な手法である概念 uni-gram と bi-gram を利用していることである。これらの手法では、語義の識別の手掛かりとなる単語が2単語以上離れた位置に存在している場合、その手掛かりを発見することができない。もう1つの問題は、概念 uni-gram と bi-gram を評価データから構築していることである。教師なし学習手法では、概念と単語間の共起の強さが必要である。しかし、その共起の強さを算出するために評価データを使用した場合、評価データにあった共起の強さが得られるため、実際の評価時に有利に働く可能性がある。

第4章 Yarowsky の手法

4.1 概要

1つの文中において、近くに存在する単語同士は関連性が深いことを利用し、大量の文を解析することで、周囲の単語から想起される単語を計算することができる。この想起される単語を語義として、語義曖昧性解消に利用する。単語の持つ語義を概念とし、概念は複数の語義を持つ。このとき、単語の持つ語義同士は同じ概念に属さない。

共起の強さは、曖昧単語の持つそれぞれの語義の属する概念と文中に存在する共起単語間で計算する。

4.2 共起の強さの計算

共起の強さの計算例を以下に示す.

文「私/は/茨城大学/へ/ゆく/.」が存在し、「行く」の語義曖昧を解消するために「行く」の持つ語義と共起単語間の共起の強さを計算する. まず,「ゆく」は語義「語義1」と「語義2」の2つを持ち,それらの属する概念はそれぞれ「A」と「B」とであると仮定する. また,文中の共起単語は「私」と「茨城大学」とであると仮定する. 図 4.1 に仮定を図示する.

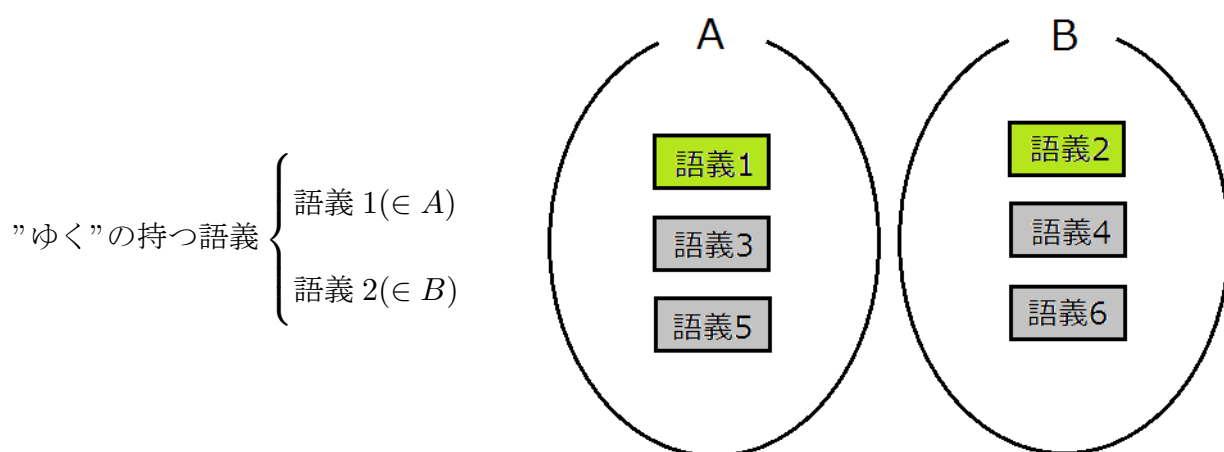


図 4.1: ”ゆく”の持つ語義と語義の属するクラス

これらの仮定から, 行くの語義と共起単語間の共起の強さは図 4.2 のように全ての組み合わせで計算される. また, 共起の強さを加算する値は自分で定義する.

共起の強さ (A, ”私”)++
共起の強さ (A, ”茨城大学”)++
共起の強さ (B, ”私”)++
共起の強さ (B, ”茨城大学”)++

図 4.2: 共起の強さ計算

4.3 語義曖昧解消

計算された共起の強さをを用いて語義曖昧解消を行う例を以下に示す。

文「私/は/この/方法/で/ゆく/こと/に/決定/した/.」が存在し、曖昧単語「ゆく」の語義曖昧性解消を行う。まず、文中に存在する共起単語は「私」、「方法」、「決定」、「した」と仮定する。「ゆく」は語義「語義1」と「語義2」の2つを持ち、それらの属する概念はそれぞれ「A」と「B」と仮定する。

曖昧単語「ゆく」の持つ語義の属する概念と共起単語との共起の強さを合計し、最も共起しやすい概念を決定する。最も共起しやすい概念が持つ曖昧単語の語義が真の語義である。表 4.1 には概念と共起単語間の共起の強さの合計の計算例を示す。この例の語義曖昧性解消の結果は、合計の大きい概念 A 中の曖昧単語「ゆく」の語義である「語義1」が真の語義となる。

表 4.1: 共起の強さの合計

概念 A	概念 B
共起の強さ (A, "私") 0.5	共起の強さ (B, "私") 0.5
共起の強さ (A, "方法") 0.5	共起の強さ (B, "した") 0.5
共起の強さ (A, "決定") 0.5	
共起の強さ (A, "した") 0.5	
合計 2.0	合計 1.0

第5章 実験

5.1 概要

本論文では、概念と単語間で計算した共起の強さから Yarowsky の手法を用いて教師なし語義曖昧性解消を行う。語義が属する概念は、新納らが作成した語義を見出しとした概念辞書を利用して決定する。

本論文では、曖昧単語を予め指定しておき、曖昧単語が持つ語義の属する概念と共起単語間の共起の強さを計算する。概念と共起単語間の共起の強さを計算するときに用いるコーパスは、「'95年版毎日新聞記事」とする。概念と共起単語間の共起の強さは、1文内の共起単語を対象として計算する。

概念と単語間の共起の強さから Yarowsky の手法を用いて語義曖昧性解消を行う。Yarowsky の手法によって得られた語義と評価データの真の語義を比較して、語義曖昧性解消の正答率を得る。本実験では、評価データとして SemEval-2 の日本語辞書タスクのテスト用例を用いる。

5.2 共起の強さ

本実験で用いる共起の強さを計算するために用いるコーパスは mecab で形態素解析されている。

本論文において、共起単語として相応しい単語の品詞は、「名詞」と「動詞」とした。さらに、「動詞」について品詞細分類 1 が「自立」のときのみとした。これは、「動詞」の品詞細分類 1 として出現する「接尾」と「非自立」が共起単語の品詞として相応しくないと考えたためである。

また、単語 A が持つ語義の属する概念と単語間の共起の強さを計算する際に加算される値は、単語 A が持つ語義数の逆数とする。

以上の設定より、共起の強さを計算する操作は以下の操作を順番に行う。

1. コーパスから 1 文取り出す
2. 文から共起単語を取り出す
3. 共起単語に曖昧単語の持つ語義の属する概念を持つ単語が存在する場合、4へ。存在しない場合、1に戻る
4. その単語の持つ概念とそれ以外の共起単語間で共起の強さを計算する
この時に、この概念の出現頻度も記録する(後述)
5. 1に戻る

5.3 語義曖昧性解消

語義曖昧性解消を行うために、曖昧単語が含まれる文から共起単語を取り出す。本論文では、共起単語は品詞が「名詞」の場合は、細分類に関わらず共起単語とし、「動詞」の場合は、細分類が「一般」のときのみとした。これも、共起の強さの計算時と同様に「動詞」の品詞細分類 1 として出現する「非自立可能」が共起単語の品詞として相応しくないと考えたためである。

また、Yarowsky の手法において、曖昧単語の持つ語義の概念と共起単語間の共起の強さの合計が最も大きい概念の語義が真の語義となるが、共起の強さの合計が等しい概念が複数現れた時の処理を本論文では以下のように定義する。これに伴い共起の強さ計算時に概念の出現頻度を記録する操作が必要となる。

1. 共起の強さが等しい場合、概念の出現頻度が高い方を選択する
2. 出現頻度が等しい場合、概念の名前昇順で早い方を選択する

5.4 評価データ

本論文では，評価データとして SemEval-2 の日本語辞書タスクのテスト用例を用いる．この評価データには，表 5.1 に示す 50 単語と，単語ごとに 50 個の例文が存在する．本論文では，表 5.1 の 50 単語を曖昧単語とし，曖昧単語の持つ語義の属する概念と単語間の共起の強さを計算する．

表 5.1: 曖昧単語

相手	会う	上げる	与える	見る
生きる	意味	入れる	大きい	持つ
教える	可能	考える	関係	求める
技術	経済	現場	子供	もの
時間	市場	社会	情報	やる
進める	する	高い	出す	良い
立つ	強い	手	出る	他
電話	取る	乗る	場合	前
入る	はじめ	始める	場所	見える
早い	一	開く	文化	認める

5.5 結果

評価データの 50 個の曖昧単語を 50 個の用例文を用いて語義曖昧性解消を行った時の全曖昧単語の平均正答率は，46.08 % となった．また，各曖昧単語の平均正答率を表 5.2 に示す．

表 5.2: 単語の正答率

単語	本論文	Std-0	Std-1
相手	0.46	0.78	0.80
会う	0.50	0.88	0.92
上げる	0.18	0.44	0.52
与える	0.10	0.76	0.70
生きる	0.30	0.94	0.94
意味	0.28	0.48	0.44
入れる	0.24	0.74	0.74
大きい	0.94	0.94	0.94
教える	0.24	0.36	0.52
可能	0.56	0.68	0.64
考える	0.96	0.98	0.98
関係	0.48	0.96	0.96
技術	0.38	0.84	0.82
経済	0.98	0.98	0.98
現場	0.80	0.74	0.76
子供	0.54	0.60	0.62
時間	0.12	0.86	0.84
市場	0.44	0.52	0.56
社会	0.14	0.86	0.86
情報	0.18	0.86	0.84
進める	0.66	0.92	0.92
する	0.00	0.64	0.72
高い	0.82	0.86	0.88
出す	0.10	0.40	0.50
立つ	0.16	0.52	0.50

単語	本論文	Std-0	Std-1
強い	0.86	0.92	0.90
手	0.08	0.78	0.78
出る	0.30	0.52	0.52
電話	0.72	0.84	0.78
取る	0.02	0.26	0.28
乗る	0.40	0.78	0.78
場合	0.72	0.84	0.84
入る	0.40	0.54	0.56
はじめ	0.68	0.88	0.88
始める	0.76	0.88	0.86
場所	0.32	0.90	0.96
早い	0.54	0.70	0.70
一	0.92	0.92	0.90
開く	0.54	0.78	0.84
文化	0.96	0.98	0.98
他	1.00	1.00	1.00
前	0.68	0.76	0.76
見える	0.46	0.68	0.70
認める	0.02	0.76	0.82
見る	0.10	0.78	0.78
持つ	0.02	0.78	0.80
求める	0.40	0.64	0.76
もの	0.88	0.88	0.88
やる	0.02	0.96	0.96
良い	0.68	0.56	0.54
平均	0.4608	0.757	0.769

第6章 考察

6.1 他手法との比較

6.1.1 baseline システム

SemEval-2 のコンペで baseline とされた SVM のシステムと比較する. このシステムは, 表 6.1 の 20 種の素性する.

表 6.1: 素性

e1= 二つ前の単語,	e2= 二つ前の品詞,	e3= その細分類,
e4= 一つ前の単語,	e5= 一つ前の品詞.	e6= その細分類,
e7= 問題の単語,	e8= 問題の単語の品詞,	e9= その細分類,
e10= ひとつ後の単語,	e11= ひとつ後の品詞,	e12= その細分類,
e13= 二つ後の単語,	e14= 二つ後の品詞,	e15= その細分類,
e16= 係り受け,	e17= ふたつ前の分類語彙表の値 (5 桁),	
e18= ひとつ前の分類語彙表の値 (5 桁),		
e19= ひとつ後の分類語彙表の値 (5 桁),		
e20= ふたつ後の分類語彙表の値 (5 桁)		

本来の baseline のシステムでは分類語彙表 ID の 4 桁と 5 桁を同時に使う形になっていたが, ここでのシステムでは 5 桁のみとした. また, 一般に一つの単語に対しては複数の分類語彙表 ID が存在するので, e17,e18,e19,e20 に対する素性は複数になる場合もある. このシステムの正解率を 14 ページ表 5.2 に示す. Std-1 は素性として表 6.1 の 20 種類全ての素性を用いた結果であり, Std-0 は素性としてシソーラスの情報 (e17, e18, e19, e20) を除いた 16 種類の素性を用いた結果である.

また, 新納の手法では概念 uni-gram だけを用いた場合, 正答率は 0.6681 となり, uni-gram と bi-gram を用いた場合, 0.6746 である.

表 5.2 より，本論文の正答率は，baseline の結果である Std-0，Std-1 の正答率よりも約 30 % 低い。また，新納の手法の問題点を解決した結果，正答率が約 20 % 低くなった。

6.2 新納の手法の問題解決

新納の手法において 2 つ問題があった。1 つの問題は，語義識別の手掛かりが発見できないことである。これは，概念辞書を用いてコーパスに対して概念 uni-gram と bi-gram を行い，単語に概念を割り当てるため，2 単語以上離れたところに存在する単語は無視することになるためである。本論文では，共起の強さを測るため，1 文内の共起単語に対して計算を行っている。よって，1 文内に存在する語義識別の手掛かりを見逃すことがなくなる。

もう 1 つの問題は，評価データを共起の強さを測るために使用しており，評価時にこの評価データにあった共起の強さが有利に働く可能性があることだ。本論文では，共起の強さを計算するためのデータと評価データは別のデータにしたため，公正な共起の強さを得ることができ，正当に評価できると考えられる。

第7章 結論

本論文では、新納らの手法において問題であった点を解決した教師なし語義曖昧性解消を行った。結果、正答率は低下し、問題点が3つ発見された。1つの問題点は、評価時に例文内の共起単語から概念との共起の強さが抽出できず、語義識別ができていないことがある。これは、共起の強さを計算するためのコーパスにおいて、概念と単語の両方が存在する文が無いことが原因である。解決策としては、共起の強さを計算するためのコーパスの量を増加させることがあげられる。

もう1つの問題点は、概念辞書に登録されている単語の語義が1つしかなく、評価時に語義の選択ができないため、実質語義曖昧性解消になっていないことがある。これは、概念辞書作成時に単語の語義が1種類しかコーパスに存在しないことが原因である。解決策としては、概念辞書の作成時に用いる語義タグ付きコーパスに語義が1つしかない単語の他の語義が付いた文を追加することがあげられる。

最後の問題点は、正答率が極端に低い単語「する」の正答率 0.00 や「認める」の正答率 0.02 が出現することである。これらの単語の正答率を改善することで大幅に平均正答率を改善することができる。

今後は以上3つの問題点を解決しながら正答率向上に努力する。

謝辞

本研究に際して，多くのご指導を頂きました指導教員の新納浩幸教授に心より感謝致します。また，日常の議論を通じて多くの知識や示唆を頂いた新納研究室の皆様にも感謝致します。

参考文献

- [1] Manabu Okumura, Kiyooki Shirai, Kanako Komiya, and Hikaru Yokono. SemEval-2010 Task: Japanese WSD. In *The 5th International Workshop on Semantic Evaluation*, pp. 69–74, 2010.
- [2] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *ACL-95*, pp. 189–196, 1995.
- [3] 小野寺喜行. 領域間距離を利用した能動学習による語義曖昧性解消の領域適応. 平成 25 年度茨城大学工学部情報工学科卒業研究論文, p. 8, 2014.
- [4] 新納浩幸, 古宮嘉那子, 佐々木稔. all-words wsd のための概念辞書の自動作成. 情報処理学会自然言語処理研究会, pp. NL-224-13, 2015.