

複数の知識源を利用した語義曖昧性解消手法のエラー分析

白井 清昭

北陸先端科学技術大学院大学 情報科学研究科

kshirai@jaist.ac.jp

1 はじめに

エラー分析はシステムの誤りの要因を分析する作業である。現状のシステムの問題点を分析し、それに対する解決策を講ずることは、科学技術の発展に欠かせない。本論文では、語義曖昧性解消 (Word Sense Disambiguation; WSD) を対象としたエラー分析の一つの事例研究について報告する。分析の対象とするのは語義タグ付きコーパスと機械可読辞書を知識源として語義を判別するシステムである。現在主流となっている教師あり機械学習に基づく手法だけでなく、機械可読辞書を用いた手法もエラー分析の対象とすることで、現状の WSD 技術に関するより多くの問題点を洗い出すことを狙う。また、1つのシステムの問題点を明らかにするだけでなく、前処理として用いる自然言語処理技術、利用する知識、WSD のタスク設定などについての問題も分析し、WSD 技術の今後の進展についての展望を得ることを目的とする。なお、本研究は Project Next NLP の語義曖昧性解消タスク [11] の活動の一貫として行われた。

2 関連研究

WSD に関するエラー分析は、論文で提案されているシステムの評価実験の中で行われていることが多い。例えば、Moldovan and Novischi は、WordNet の語釈文中の単語の語義を自動的にタグ付けするためのルールベースの WSD 手法を提案し、事例を紹介しながらルール毎に誤りの要因を分析している [5]。Boyd-Graber et al. は、WordNet 上のランダムウォークを単語の生成過程として組み込んだ Latent Dirichlet Allocation (LDA) [1] の拡張モデルを提案している [2]。与えられた文書に対し、提案する確率モデルの事後分布を推定することで文書内の全ての語の語義ならびに文書のトピックを決定する。エラー分析では、どのようなトピックの文書にも現われる高頻度語の語義の判別に失敗したり、低頻度だが WordNet において多くの synset とつながりを持つ語義が誤って選ばれる場合が多かったと報告している。エラー分析から得られた知見を基に新しい WSD 手法を提案した事例として藤

井らの研究 [6] がある。彼らは、WSD の評価型ワークショップである SENSEVAL-2 日本語辞書タスク [12] の対象単語のうち、人間は容易に判別できるが計算機では判別が難しい 6 単語を対象に語義判別の誤りの原因を調査した。その結果、遠方単語 (WSD の対象語からの距離が遠い語) の影響は少なく、未知語 (テスト文では対象語の近くに出現するが、訓練データにおいては出現しない語) の影響の方が大きいことを明らかにした。これを踏まえ、ウェブ検索のヒット件数を基に単語間の類似度を測り、これを WSD に応用する手法を提案している。ただし、ほとんどの先行研究におけるエラー分析は提案システムの検証だけに留まっている。WSD の問題点を包括的に調査すること目的とし、エラー分析を主眼に置いた研究は行われていない。

白井は、SENSEVAL-2 日本語辞書タスクの参加システムに関する考察を行っている [12]。複数の参加システムのうち 1つのシステムのみで不正解だった事例を紹介し、このような事例を調べることでシステムの欠点を探ることができると述べている。また、語義のタグ付け作業者の一致率が高いのにも関わらず参加システムの正解率が低い単語の例を挙げ、このような対象単語に対するシステムの出力を調べることによってシステムを改善できると述べている。しかし、エラー分析の方向性を示してはいるが、詳細なエラー分析に基づく WSD システムの改善方法の提案は行っていない。

3 分析手順

3.1 WSD システム

エラー分析の対象とした WSD システムは文献 [13, 14] の手法を一部改変したものである。ここではその概略について述べる。本システムは、(1) 機械学習による分類器、(2) 辞書の用例に基づく分類器、(3) 辞書の文法的制約に基づく分類器の 3つの分類器を組み合わせたものである。(1) は語義タグ付きコーパスである SemEval-2010 日本語 WSD タスクのデータ [10] を、(2) と (3) は機械可読辞書である岩波国語辞典 [9] を知識源とする。

3.1.1 機械学習による分類器

語義タグ付きコーパスから学習した Support Vector Machine (SVM) による分類器である。SVM の学習には LIBSVM¹ を利用し、カーネルは線形カーネルを用いた。学習素性は以下の通りで、いずれも WSD でよく用いられる素性である。

- コロケーション
 - 対象語の前後 1~2 語の単語、その品詞
 - 直前 2 語、直後 2 語、前後 2 語の単語の組、その品詞の組
 - 対象語自身の表記
 - 文脈内の自立語
 - 同一文に出現する自立語の基本形
 - 上記の素性の意味クラス
 - 構文素性
 - 係り先文節の主辞 *1
 - 係り元文節の主辞 *1
 - 文節内の主辞 *2
 - 格と格要素の組 (名詞+助詞) *3
 - 格と格要素の意味クラスの組 *3
 - 格と係り先用言の組 *4
- *1 は対象語が文節の主辞のとき、*2 は文節の主辞ではないとき、*3 は動詞のとき、*4 は名詞のとき、それぞれ抽出する。

上記の素性における意味クラスは日本語語彙大系 [8] における上位の分類カテゴリを用いた。また、素性を抽出するため、形態素解析には MeCab² を、文節の係り受け解析には CaboCha³ を用いた。

3.1.2 辞書の用例に基づく分類器

本システムにおける語義の定義は岩波国語辞典にしたがう。岩波国語辞典では定義文の他に用例が記載されている。この分類器は、対象語を含む文 (テスト文) と辞書の用例の類似度を計算し、最も似ている用例に対応する語義を選択する。ただし、類似度が十分に高い用例が見つからないときは、この分類器は語義を選択しない。

テスト文 t と辞書の用例 e の類似度は式 (1) のように定義する。

$$Sim(t, e) = \max_r sim_w(w_t^r, w_e^r) \quad (1)$$

¹<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

²<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

³<https://code.google.com/p/cabocho/>

w_t^r, w_e^r はそれぞれ文 t, e において対象語と r の関係にある語を表わす。関係 r は以下のいずれかとする。

- 格関係 (用言と格要素の関係)
- 助詞「の」を介した係り受け関係
- 複合名詞内で連続して出現する名詞間の関係
- 形容詞とそれに係る名詞の関係

また、 sim_w は二単語間の類似度で、日本語語彙大系によって測る。

例えば、文 S1 において「始めれ」が WSD の対象語であるとする。

S1 中学校に入って本格的に勉強を*始めれ*ば、

図 1 は「はじめる」の語釈文である。S1 と 41138-0-0-1 の用例「勉強を一」⁴ との類似度は、ヲ格の名詞がともに「勉強」であることから他の用例よりも高いため、41138-0-0-1 の語義が選択される。

3.1.3 辞書の文法的制約に基づく分類器

岩波国語辞典では、その語義が使われる際の文法的な制約が記述されていることがある。例えば、図 1 における《》で囲まれた記述は、41138-0-0-2 や 41138-0-0-3 の語義が出現するときに満たされるべき制約である。この分類器は、テスト文が語釈文の中の文法的な制約を満たすとき、その語義を選択する。制約を満たす語義がないとき、あるいはどの語義についても文法的制約が辞書に記述されていないときは、語義を選択しない。

例えば、文 S2 の「始め」が対象語のとき、そのひとつ前の単語が動詞であり、《動詞や「(さ)せる」「(ら)れる」の連用形に付き》という制約を満たすことから、41138-0-0-2 の語義が選択される。

S2 だって目立ち *始め* たのって、去年の終わりから

3.1.4 混合モデル

混合モデルは上述の 3 種類の分類器を組み合わせて用いる。複数の分類器が語義を選択したとき、式 (2) にしたがって最終的な出力語義 s' を選択する。

$$s' = \arg \max_{cl \in CL, s \in S} prec(cl, s) \quad (2)$$

CL は 3 種類の分類器の集合、 S は語義の集合で、 $prec(cl, s)$ は分類器 cl が語義 s を選択したときの精度を表わす。 $prec(cl, s)$ は、SVM については訓練データ上での 10 分割交差検定で、辞書の用例ならびに文法的制約を用いる分類器については SVM の訓練データを用いて算出する。

⁴岩波国語辞典の用例では傍線「一」は見出し語を表わす。

語義 ID	語釈文
41138-0-0-1	物事を新たに行う。今までの、していない状態から、する状態に移す。「勉強をー」「店をー」「今に一ぬ事ながら」(今新たに始めたというたぐいの事ではないが)。また、当人の癖と見られているしぐさ・言い方をし始める意にも言う。「ほらーぞ。いつもの小言だ」
41138-0-0-2	《動詞や「(さ)せる」「(ら)れる」の連用形に付き、接尾語的に》新たにその動作を起こす。…しだす。…ことが始まる。「本を読みー」「花が咲きー」「書かせー」
41138-0-0-3	《「…をー」の形で》↓はじめ (3)

図 1: 「はじめる」の語釈文

表 1: WSD システムの評価

	精度	再現率	F 値	適用率
MFS	0.7692	0.7692	0.7692	1.0000
SVM	0.8012	0.8012	0.8012	1.0000
EX	0.7619	0.0512	0.0960	0.0672
GR	0.7364	0.0380	0.0723	0.0512
ALL	0.8024	0.8024	0.8024	1.0000

3.2 分析データ

評価データとして、SemEval-2010 日本語 WSD タスクのテストデータを用いる。このデータは、現代日本語書き言葉均衡コーパス (BCCWJ)[16] に語義を付与したコーパスである。50 語を対象語とし、それぞれの語について 50 用例、合計 2500 の用例からなる。一方、SVM の訓練データは同じく BCCWJ の語義タグ付きデータを用いる。1 つの評価単語に対する訓練データの用例数は 49~6733 であり、その平均は 298 である。

WSD システムの評価結果を表 1 に示す。一列目は分類器で、MFS は最頻出語義を常に選択するベースライン、SVM は機械学習による分類器、EX は辞書の用例を用いた分類器、GR は文法的制約を用いた分類器、ALL は混合モデルを表わす。「適用率」は分類器が語義を選択することのできたテスト文の割合である。EX や GR は 5~6% のテスト文に対してしか語義を選択できないため、再現率も低い。SVM と混合モデルではほとんど差がないことから、今回の実験では 3 種類の分類器を併用する効果は見られなかった。しかし、機械学習に基づく手法だけでなく、他のアプローチによる WSD 手法もエラー分析の対象としたいため、今回の分析では混合モデルの出力を用いる。

混合モデルで正しい語義が選択できなかった事例の数は 494 であった。これらのうち、以下のいずれかの条件を満たす 201 事例を今回のエラー分析の対象とした。() 内は各条件に当てはまる事例数である。

1. Project Next NLP の語義曖昧性解消タスクのグループによってエラー分析 [11] した 50 事例と重

複する事例 (23)

なお、[11] ではコロケーション、係り受け、分類語彙表の意味クラスを素性とした SVM を対象にエラー分析をしており、本研究とは使用する WSD システムが異なる。

2. 辞書の用例、文法的制約に基づく分類器によって語義が選択された事例 (30)
3. 1,2 以外の事例からランダムに選ばれた事例 (148)

4 エラー分析

誤り事例を詳細に分析し、どのような要因で正しい語義が選択できなかったかを分類した。エラー分析のおおまかな進め方は以下の通りである。

- 正しい語義を判定するための根拠を探す。例えば、直前の単語、格関係、文脈内の自立語によって語義が決まると考えられる場合には、それらが根拠となる。その根拠が SVM の学習素性として表現されているかを確認する。
- 正しい語義を判定するための根拠が素性として表現されているとき、その素性が訓練データの中に出現するかを確認する。
- SVM の訓練データにおける素性と語義の共起度を Dice 係数で測る。正解語義ならびに出力語義について、共起度の高い素性を表示させ、これらを比較したり、抽出した素性に不備がないかを調べる。
- テスト文と似た文が辞書の用例にあるか、両者の類似度が適切に計算されているか、されていないときはその原因は何かを調べる。
- その他、誤りの原因と考えられるものは全て挙げる。

今回のエラー分析で発見できたエラーの要因の一覧を図 2 に示す。エラーの要因を大きく【手法の問題】【知識の問題】【前処理の問題】【データの不備】【問題設定の不備】の 5 つに分類し、それぞれをさらに細かく分類して全体を階層構造で表現した。今回のエラー分析では発見できなかったが、一般に誤りの要因とし

手法の問題	
教師あり機械学習に基づく手法の問題	
訓練データの不足	(27)[0.134]
└ 他に手がかりなし	(21)[0.104]
素性抽出が不適切	(2)[0.010]
└ 意味クラスの抽象度	(5)[0.025]
└ 助詞の取り扱い	(10)[0.050]
└ 格の交替の取り扱い	(3)[0.015]
└ 連体修飾の取り扱い	(8)[0.040]
└ システムのバグ	(3)[0.015]
有効な素性の不足	(7)[0.035]
└ トピック素性	(10)[0.050]
└ 長いコロケーション	(2)[0.010]
└ 間接的な係り受け	(3)[0.015]
└ 既存の素性の組み合わせ	(7)[0.035]
└ 文脈に出現する語の語義	(2)[0.010]
└ 語釈文と文脈の関連性	(3)[0.015]
└ 照応・省略解析	(3)[0.015]
素性のコーディングが困難	(1)[0.005]
└ 文の解釈	(20)[0.100]
└ 文脈の解釈	(18)[0.009]
学習アルゴリズムの問題	(14)[0.070]
└ 過学習	
辞書の用例に基づく手法の問題	
文間類似度の不備	
└ 類似度が低すぎる	(7)[0.035]
└ 類似度高すぎる	(20)[0.100]
└ 表層的には似ていない	(6)[0.030]
システムのバグ	(1)[0.005]
タイプブレークが不適切	(1)[0.005]
辞書の文法的制約に基づく手法の問題	
└ 文法的制約が緩い	(7)[0.035]
└ 規則の不備	(1)[0.005]
分類器の組み合わせ手法の問題	(14)[0.070]
└ 消去法	(14)[0.070]
知識の問題	
└ シソーラスの不備	(3)[0.015]
前処理の問題	
└ 形態素解析の誤り	
└ 文節の係り受け解析の誤り	(1)[0.005]
データの不備	
└ 正解語義の誤り	
└ 訓練データ	(15)[0.075]
└ テストデータ	(32)[0.159]
問題設定の不備	
└ 文脈不足	(1)[0.005]
└ 対象語が不適切	(16)[0.080]
└ 熟語・連語として扱う方が適切	(5)[0.025]
└ 人間でも判定が困難	
その他	(1)[0.005]

図 2: エラー要因の分類

て発生し得ると考えられるものも挙げた。()内の数字は該当する誤り事例の数, []は全体(201事例)に対する割合である。ひとつの誤り事例が複数の要因に分類されることがある。以下、それぞれのエラー要因について説明する。

4.1 手法の問題

4.1.1 教師あり機械学習に基づく手法の問題

使用した3つの分類器のうち、最も貢献度が高いのはSVMだったので、これについては詳細に分析した。以下、それぞれのエラー要因について述べる。

4.1.1.1 訓練データの不足

訓練データにテスト文と似た事例がないために正しい語義を選ぶことができなかったとき、このタイプに分類した。ただし、現実的には訓練データの量が十分でない場合の方が多いので、このタイプに分類するだけではWSDの問題の解決策を探ることにはならない。そこで、このエラー要因に分類したときは他の要因も探した。ただし、テスト文に類似した事例が訓練データにないと語義を判別しようがない場合は【他に手がかりなし】とした。その多くは定型的な言い回しで語義が決まる事例である。例えば、文S3の「被害に*逢う*」は決まり文句に近く、この文が訓練データにないと「あう」が166-0-2-3[物事に会おう]という意味であると判断することは難しい。

S3 子供が被害に*遭う*事件が相次いでいる。

二格の格要素「被害」を意味クラスによって抽象化することも検討したが、166-0-2-3[物事に会おう]という語義の二格の選択制約を適切に表わすような意味クラスは発見できなかった。

4.1.1.2 素性抽出が不適切

テスト文からの素性の抽出に問題が見つかったとき、このタイプに分類した。さらに、問題の種類に応じて以下の5つに細分類している。なお、図2で【素性抽出が不適切】に分類した事例が2つあると示されているが、これらは下記の5つの細分類に当てはまらない例外的な事例である。

【意味クラスの抽象度】

素性として使用した意味クラスの抽象度が高く、語義の判別に貢献しなかった事例である。例として文S4を挙げる。

S4 マグロのほうは売れば、ガソリン代が*で*て、

訓練データには「お金が出る」「賞金が出る」という用例があり、S4がこれらと類似していることがわか

れば正しい語義を選択できる。このとき「金」「賞金」「(ガソリン)代」が〈お金〉のような意味クラスで表現できればよい。ところが、WSDシステムでは「代」を1155(制度)⁵という抽象度の高い意味クラスに変換していた。すなわち、素性として使用する意味クラスの抽象度の設定が適切ではなかった。

【助詞の取り扱い】【格の交替の取り扱い】【連体修飾の取り扱い】

表現の正規化をした上で素性を抽出すべき事例である。例を表2に挙げる。【助詞の取り扱い】の例では、「でも」のような複合助詞を「で」に直してから素性を抽出すべきである。【格の交替の取り扱い】の例では、受身形、使役形の動詞については格の交替を考慮すべきである。【連体修飾の取り扱い】の例では、内の関係の連体修飾の場合は格を補った上で素性を抽出すべきである。

表2: 表現の正規化が必要な例

【助詞の取り扱い】	病院 <u>でも</u> *診*て貰えない → 病院 <u>で</u> *診*て貰えない
【格の交替の取り扱い】	退去勧告が*出さ*れている。 → 退去勧告を*出す*。
【連体修飾の取り扱い】	*出る*熱を → 熱が*出る*

【システムのバグ】

WSDシステムの実装上のバグがいくつか見つかった。バグを発見することもエラー分析の意義のひとつである。

4.1.1.3 有効な素性の不足

WSDの手がかりとなる情報が素性として利用されていない場合、このタイプに分類した。また、その情報の種類に応じて以下のように細分類した。

【トピック素性】

対象語の語義と文のトピックとに強い関連性があるとき、トピックは語義を決める手がかりとなる。例えば、以下のS5はスポーツに関連する文であり、「相手」が単なる相棒や仲間ではなく117-0-0-3[自分と対抗して物事を争う人]の語義を持つとわかる。

S5 新聞は「体重六十六キロの日本人が七百三十二キロを破る」とか「六十六キロが五百七十五秒で*相手*をすべて倒した」と書き立てた。フランス柔道の

本論文のシステムではこのようなトピック素性は用いていないが、LDAの潜在的トピックを素性として利用した先行研究[3]がある。

⁵1155は日本語語彙大系の分類コードである。

【長いコロケーション】

本論文ではコロケーションの素性として2単語までの組み合わせしか利用していないが、3語以上の組み合わせがWSDに有効な素性となる事例である。例を以下に示す。

S6 脂肪を落とすという*意味*なら二の腕のみを

S6の対象語の「意味」には2843-0-0-1[その言葉の表す内容]と2843-0-0-2[表現や行為の意図・動機]という語義がある。本論文のシステムではテスト文から「と/いう/(意味)」というコロケーションが素性として抽出されるが、これは訓練データでは両方の語義の周辺に出現する。しかし、「と/いう/(意味)/なら」というコロケーションは2843-0-0-2の意味としか共起しない。このように3語のコロケーションを素性としてすることで正しい語義が選択できる可能性がある。

【間接的な係り受け】

本論文のシステムにおける構文素性は対象語との直接の係り受け関係のみであるが、間接的な係り受け関係がWSDに有効なときがある。例えば、以下の文S7では、「親が子供を監視する」という文脈から「子供」の語義が17877-0-0-1[幼い子, 児童]ではなく17877-0-0-2[自分のもうけた子]とわかる。

S7 あとは今少子化で親が*子供*ばかりを監視し、

このとき、「親」と「子供」に直接の係り受け関係はないが、「親が」と「子供を」の係り先が「監視する」であり、「親」と「子供」は同じ用言の格要素であるという間接的な関係がある。つまり、対象語の係り先の用言に係る他の格要素(この場合は「親」)を素性としてすることで正しい語義が選択できる可能性がある。

【既存の素性の組み合わせ】

既存の素性の組み合わせが有効と考えられる事例である。

S8 経験=古き*良き*時代を、回答することで

この例における対象語「よい」には52935-0-0-1[物事が他よりまさった状態にある]、52935-0-0-2[正しい]、52935-0-0-3[好ましい, 適している]などの語義がある。正解語義は52935-0-0-3である。対象語自身の表記の素性「良き」や直後の単語の表記の素性「時代」はその語義とも共起するが、これらの組み合わせ「良き時代」は52935-0-0-3の語義としか共起しないと考えられる。

【文脈に出現する語の語義】

文脈に出現する語の語義の情報が必要な事例である。

以下の文 S9 において、対象語「挙げ」の意味は、ヲ格の格要素「点」がこの文のように抽象的な意味を持つか、あるいはスポーツの「点」の意味を持つかによって決まると考えられる。

S9 前年比5%減になっている点を*挙げ*た。

【語釈文と文脈の関連性】

語釈文中の単語と対象語の文脈に出現する語の連想関係が WSD の手がかりとなる事例である。以下の文 S10 において、正解語義の語釈文 31640-0-0-4[水中などに直立して、顔が没しない]の中の「水中」と文脈に出現する「浅瀬」の間に連想関係があることが正しい語義を選択する手がかりとなる。

S10 怖くなって足の*立つ*浅瀬に引き返してくる

【照応・省略解析】

照応解析によって代名詞の参照先を決めたり、省略解析によって省略された要素を補完する処理が必要な事例である。例えば、文 S11 の対象語「開い」の語義を決定するためには、省略解析によってヲ格の格要素がひとつ前の文の「封筒」であることを同定する必要がある。

S11 請求書ご案内と請求書、各一通を封筒に入れる時どちらを先に入れます？ *開い*たときに「請求書ご案内」が上に来るように入れます。。

この他、上記の細分類に入れなかったものとして、固有表現抽出の結果や one sense per discourse[15] の制約などを素性とすると改善が見込める少数の事例が見つかった。

4.1.1.4 素性のコーディングが困難

語義を決める手がかりは発見できたが、これを抽出するには高度な言語処理や推論を必要とし、機械学習の素性として表現することが難しい場合、このタイプに分類した。テスト文のみの深い解釈が必要な場合の【文の解釈】と、対象語を含む文章全体の解釈が必要な場合の【文脈の解釈】とに細分した。前者の例として S12、後者の例として S13 を挙げる。

S12 自分も休みの日くらいになりたいのに、*子供*の面倒を凄く見てくれます。

S13 調理道具がそろってなくても、炊飯器さえあれば色々作れる。初めて自炊する人や男性にも楽しんでもらおうと*考え*ました。ご飯物や肉じゃがといった定番料理、おやつなど百八品を取めた本は、

S12 の対象語「子供」には 17877-0-0-1[幼い子、児童]と 17877-0-0-2[自分のもうけた子]の語義があるが、

『子供の面倒を見る』という文ではどちらの語義もあろう。『自分も休みの日くらいになりたい』という節の意味を解釈すると「子供」が自分の子を指すことがわかる。S13 の対象語「考える」には 9590-0-0-1[あれやこれやと思いをめぐらす]と 9590-0-0-2[新たなものをくふうする]の意味がある。正解語義は 9590-0-0-2 であるが、この文を書いている人が料理方法を創意工夫しているという文脈の解釈が必要である。

4.1.1.5 学習アルゴリズムの問題

WSD に必要な素性は抽出されていて、類似用例も訓練データに存在するが、SVM で学習された分類器では正解を選択できなかった事例である。テスト文から抽出された素性のうち、語義との共起度が高い上位の素性を調べ、それらの素性がシステムの出力語義よりも正解語義との共起度の方が高いのにも関わらず、また正解語義が訓練データにおける最頻出語義であるのにも関わらず、システムが正解語義を選択できなかった場合はこのタイプに分類した。これは素性の設計には問題がなく、学習アルゴリズムに問題があると考えられる。SVM 以外の機械学習アルゴリズムでは正しく解ける可能性がある。

4.1.1.6 過学習

今回のエラー分析では該当する事例は見つからなかったが、訓練データに特殊な用例が含まれていて、そのために判定を誤った場合を想定している。

4.1.2 辞書の用例に基づく手法の問題

テスト文と辞書中の用例の類似度を測って語義を選択する分類器において、両者の類似度を適切に計算できなかったことが誤りの原因であったとき、【文間類似度の不備】に分類した。これらのうち、【類似度が低すぎる】は、テスト文と用例との類似度を低く見積り過ぎたために正しい語義を選択できなかった事例である。以下に例を挙げる。

S14 ようやく実用化のめどが*立つ*た。

正解語義の用例に「目算がたつ」があり、テスト文と似ている。「目算」と「めど」はともに日本語語彙大系に含まれるが、これらの類似度が十分高くなかった。

逆に【類似度が高すぎる】は不正解の語義の用例とテスト文の類似度を高く見積り過ぎたために不正解の語義を選択した事例である。以下に例を挙げる。

S15 女は両手に皿を*持つ*てキッチンから出てきて、ひとつをぼくの前に、ひとつを自分の席に置く。

S15 は不正解の語義の用例「新入生の組をもつ」との類似度が高く計算された。「皿」「組」はともに日本語

語彙大系の 2595〈単位〉という意味クラスに属するためであった。このような〈単位〉を表わす語の類似度が高く推定されることによって不適切な用例との類似度が高く算出された事例が多く見られた。

【表層的に似ていない】は、正解語義の用例の中にテスト文と意味的に似ているものが存在するが、表層的には似ていないため、類似度の計算が難しい事例である。以下に例を挙げる。

S16 学習会やミーティングの*持ち*方はさまざまです

S16 の正解語義には「会議をもつ」という用例があるが、テスト文の「学習会の持ち方」や「ミーティングの持ち方」と表層的には似ていないため、現状のシステムでは類似度を測れない。

辞書の用例に基づく分類器は、類似度が最大の用例が複数見つかったとき、SVM の訓練データにおける頻度が大きい方の語義を選ぶ。これをタイプブレークと呼ぶ。【タイプブレークが不適切】とは、正解語義の用例と不正解語義の用例の類似度が同じだが、タイプブレークによって正解語義が選ばれなかった事例である。

4.1.3 辞書の文法的制約に基づく手法の問題

【文法的制約が緩い】とは、辞書に記載された文法的制約が語義を絞り込めるほど厳しくないために判定を誤った事例である。例えば、岩波国語辞典の「する」の語義には以下のような文法的制約が記述されている。

27236-0-0-2	《「…とした」の形で》その状態である。「青い目をした女の子」
-------------	--------------------------------

しかし、「とした」という形で出現する「する」が常にこの語義を持つわけではない。つまり《「…とした」の形で》という制約は語義を1つに絞り込めるほど厳しくない。テスト文がこの条件にマッチしたときに不正解の語義が選択された事例があった。

【規則の不備】とは分類器の実装上の問題である。同じく「する」の語義 27236-0-0-6 に《「に」「と」を受けて》という文法的制約がある。この制約を満たすテスト文があり、制約を正しくチェックできれば正解語義を選択できた。しかし、現在のシステムは、パターンマッチによって辞書中の《》内の記述を文法的制約をチェックする規則に変換した上で、テスト文が制約を満たすかを判定している。しかし、上記の《～を受けて》という表現にマッチするパターンを用意していなかったため、この制約を分類器に組み込むことができなかった。

4.1.4 分類器の組み合わせ手法の問題

複数の分類器が語義を出力し、かつその中の一つは正解語義を選択していたのにも関わらず、3.1.4 で述べた手法で最終的な語義を決める際に正解以外の語義が選ばれた事例である。本論文のシステムでは分類器の出力の信頼度を式 (2) の *prec* によって評価することで最終的な語義を決めているが、分類器の組み合わせ手法を改良することで正しい語義が選択できる可能性がある。

4.1.5 消去法

正しい語義を人間が判定する際、該当しない語義を除外していくことで正解の語義がわかるとき、【消去法】というタイプに分類した。例えば、以下の文 S17 の「市場」の語義を判定する場合を考える。

S17 柔軟に対応しているのに、日本では現在、禁止されている。この結果、かえって医師の処方を経ないで入手できる*市場*が生じている。医療保険をめぐる抜本改革論議が進行中だが、法律の形式論理や歴史的事情、政治的な利害調整...

「市場」の語義は 21128-0-0-1[いちば], 21128-0-0-2[売行き先。「市場が広い」], 21128-0-0-3[売手と買手とが規則的に出会って取引を行う組織]の3つであるが、この文における「市場」は魚や野菜を売る場(21128-0-0-1)でもないし、株を取り引きする場(21128-0-0-3)でもないことから 21128-0-0-2 の語義とわかる。語義 [2] の「市場」は語義 [1] や [3] と比べると広い意味を持つ。したがって、この語義の文脈は様々であり、語義 [2] と共起する素性も多様であると予想できる。このとき、テスト文の素性と訓練データの文の素性に重なりが小さく、語義 [2] を正解語義と判定することは難しいとも予想できる。一方、語義 [1] では「青空」「鮮魚」「野菜」など、語義 [3] では「証券」「株式」など、比較的限られた語が出現するため、判定が容易と考えられる。したがって、対象語が語義 [2] を持つことを直接判定するよりも、語義 [1] や [3] に該当しないと判定した方が容易である。

対象語が該当しない語義を順に選ぶことで曖昧性を解消する手法をここでは「消去法」と呼ぶ。消去法は従来の WSD の枠組とは異なるアプローチである。ただし、消去法で解きやすい対象語と解きにくい対象語があると思われる。消去法の実現には、どのような対象語に対して消去法が有効に働くかを判別しなければならないなど、解決すべき課題も多いが、検討の価値はある。

4.2 知識の問題

【知識の問題】は、WSDシステムがシソーラスなどの外部知識を使用しているとき、その外部知識に問題があるために判定を誤る事例である。本論文のWSDシステムでは外部知識としてシソーラスしか用いていないので、この下位の分類は【シソーラスの不備】のみとした。例えば、以下の文 S18 において、二格の格要素の意味クラスが 2055(出来事)であるという素性が有効と考えられる。実際、訓練データにこのような素性は3回出現した。しかし、日本語語彙大系では、「置き引き」は 2055(出来事)という意味クラスの下位語ではなかった。

S18 三ヶ月ほど前に彼が置き引きに*会い*、

4.3 前処理の問題

【前処理の問題】は、形態素解析や文節の係り受け解析などの前処理における誤りが原因となる事例である。下位分類は【形態素解析の誤り】と【文節の係り受け解析の誤り】としているが、この分類も前処理としてどのようなツールを使うかに依存する。今回の分析では、文節の係り受け解析の誤りが1例見つかった。

4.4 データの不備

【データの不備】はWSDシステムの評価および開発に用いるデータの不備が原因と考えられるものを分類した。今回のエラー分析では、語義タグ付きコーパスに付与された語義が誤りである事例が主に見つかったので、【正解語義の誤り】という下位分類を設け、さらにこの下を訓練データの正解語義の誤り(【訓練データ】)とテストデータの正解語義の誤り(【テストデータ】)とに細分した。【訓練データ】の誤りは4.2項の【知識の問題】の下位に分類するという考え方もあるが、語義タグ付きコーパスの不備という点では【テストデータ】と共通しているので、ともに【データの不備】の下位に分類した。

今回のエラー分析ではテストデータの正解語義の違いが多く見つかったが、正解語義の判定は人によって判断が異なることもある。分類された誤り事例の一部は4.5項で後述する【人間でも判定が困難】に該当する事例である可能性もある。

4.5 問題設定の不備

【問題設定の不備】は、そもそもWSDの問題として解くべきではないと考えられる事例を分類した。【文脈不足】は、テスト文が短かく、これを読むだけでは

人間でも正しい語義を選択できない事例である。【対象語が不適切】とは、対象語と辞書の見出し語が一致しなかった例である。例えば、「釣り上げる」という語に対して辞書の見出し語「あげる」の語義が付与されているテスト文があったが、「釣り上げる」の意味として「あげる」の語義を選ぶのは適当ではない。今回使用したデータでは辞書の見出し語を含む複合語に語義が付与されている事例がいくつかあったが、これらは全て【対象語が不適切】に分類した。【熟語・連語として扱う方が適切】は、対象語が熟語・連語の一部になっており、全体で意味を解釈した方がよいと考えられる事例である。例えば、「たつ」という対象語について、「役に*立つ*」や「鳥肌が*立つ*」を含むテスト文があった。テストデータでは「立つ」の意味を解釈して正解語義が割り当てられていたが、「役に立つ」や「鳥肌が立つ」全体を熟語とみなして意味を解釈する方が多くの場合適切であろう。【人間でも判定が困難】は、対象語の語義の定義が詳細に分類されていて、人間でも語義の区別が難しい事例である。今回のエラー分析ではこれに該当する事例はなかったが、正解語義の判断に迷った場合も多かった。

5 考察

5.1 今後の展望

本項では、エラー分析によって明らかになった現状のWSD技術の主な問題点と、今後の展望としてその対応策について述べる。WSDに関する研究は教師あり機械学習に基づく手法が主流であり、また比較的良い成果が得られていることから、機械学習手法の問題点を中心に論じる。誤り事例の数が比較的多く、特に重要な問題と考えられるのは以下の4つである。

- 【素性のコーディングが困難】 39事例(19%)
語義を決めるために高度な推論などを要し、判定の手がかりとなる情報を機械学習の素性として表現しにくい事例である。ただし、このタイプに分類した誤り事例の語義を決めるためには必ず高度な推論が必要というわけではない。例えば、一見語義を決めるのが難しいテスト文でも、類似した事例が訓練データに存在すれば、高度な推論なしで正しい語義を選択できる。
他のタイプには分類しなかったが【素性のコーディングが困難】のみに分類した事例が24個あり、これらは他に有力な手がかりがなく、現時点での自然言語処理の技術では対応が難しいと考えられる。

とはいえ、誤り事例のうち12%を占めており、軽視できない。

- **【有効な素性の不足】** 37事例 (18%)
WSDの手がかりとなる情報は発見できたが、素性として導入されていない事例である。【素性のコーディングが困難】の事例とは異なり、学習素性として導入することが可能である。ただし、学習素性を増やせばWSDの正解率が必ず向上するわけではない。過学習など、学習素性を増やしたことによる悪影響も考慮する必要がある。有効な素性を自動的に選別する素性選択の手法も導入すべきである。

- **【素性抽出が不適切】** 31事例 (15%)
格の交替、連体修飾、複合助詞の取り扱いなどによる表現の正規化の処理が必要と考えられるケースが多く見られた。例えば、述語項構造解析を前処理に加えることが対策の一つとして考えられる。ただし、前処理を増やしたとき、前処理の誤りがWSDの判定に悪影響を与える可能性があることに留意する必要がある。

一方、学習素性として用いる意味クラスの抽象度が高すぎる事例がいくつか見られたが、これを適切に設定することは難しい。例えば、動詞の語義の曖昧性を解消したいとき、動詞の格要素の名詞は有効な素性であるが、その名詞を意味クラスに置き換えることで、表記は異なるが似た意味を持つ単語を同じ素性と認識し、テスト文と訓練データの文との類似性を測ることができる。このとき、意味クラスが動詞の選択制約を表わしていると考えることができる。ところが、選択制約の厳しさは、対象語あるいは語義によって異なると考えられる。したがって、意味クラスの抽象度を一律に設定するのはおそらく適当ではない。素性として意味クラスを導入するとき、その最適な抽象度を自動的に決定する手法を検討すべきかも知れない。

- **【他に手がかりなし】** 21事例 (10%)
【訓練データの不足】の下位分類であり、訓練データに類似用例がないと語義の判別が難しい事例である。訓練データを自動的にまたは半自動的に拡充するアプローチが有望である。例えば、Fujita and Fujinoは、辞書の用例と類似した例文をコーパスから検索し、これを語義が付与されたデータとして獲得することで、訓練データを自動的に作成す

る手法を提案している [7]。また、本論文のシステムのように、辞書の語釈文や用例など、語義タグ付きコーパス以外の知識源を併用することも対策のひとつと考えられる。

5.2 WSDのエラー分析の方法論

現在までに様々なWSDの手法が提案され、また今後も多くの手法が考案されるであろう。様々なWSDシステムのエラー分析の結果をまとめて新たな知見を得るためには、エラー要因の分類が統一されていること、あるいは標準のエラー要因の分類が存在することが望ましい。しかしながら、どのようなWSDシステムにも適用できる標準的なエラー要因の分類を定義することは困難である。図2に示した体系のうち、【手法の問題】の下の分類は、分析の対象とするWSDシステムに応じて設定せざるを得ない。しかしながら、【教師あり機械学習に基づく手法の問題】の下の分類、【訓練データの不足】【素性抽出が不適切】【有効な素性の不足】【素性のコーディングが困難】については、機械学習に基づくWSDシステムのエラー分析に共通して使えるかも知れない。また、【知識の問題】【前処理の問題】【データの不備】【問題設定の不備】については、どのようなシステムのエラー分析にも使えるという意味で標準的な分類であると言える。

最後に語義の定義とエラー分析の意義について考察したい。今回のエラー分析では、岩波国語辞典にしたがって語義を定義しているが、一般には語義は様々な方法で定義される。筆者は、WSDにおける語義の定義はアプリケーションによって決めるべきであると考えられる。例えば、WSDを機械翻訳のために使う場合には、語義は目標言語の単語で定義すべきであるし、情報検索のために使う場合には、テキストのトピックと関連する比較的粗い語義を定義すべきであろう。本論文のWSDシステムで岩波国語辞典によって語義を定義しているのは、先行研究 [14] では辞書引きを行う読解支援システムで用いることを前提にWSDシステムを構築したためである。ここで、WSDのエラー分析から得られた知見は語義の定義が異なる場合でも正しいかという疑問が生じる。例えば、岩波国語辞典では比較的細かく語義が分類されており、人間でも判断に迷うときがある。このような微妙な語義の違いを判別できずに間違えた事例でも、別の語義の定義の下では簡単に識別できるということがありうる。厳密には、語義の定義が異なれば、その都度エラー分析をしなければならないだろう。しかしながら、エラー分析

による定量的な知見, すなわちどのタイプのエラー要因が多いかという点については語義の定義や使用する WSD 手法の違いに依存するが, 定性的な知見, すなわちどのようなタイプの誤り要因が存在するかといった知見は語義の定義や手法の違いによらずある程度共通していると考えられる. 今回のエラー分析で洗い出されたエラー要因も, 語義の定義が変わったとしてもやはり誤りを発生する要因であると言えるだろう.

6 おわりに

本稿では WSD に関するエラー分析の事例研究について報告した. エラー分析は時間を要する大変な作業ではあるが, 4, 5 節で述べたような多くの知見が得られたことから, その重要性を再認識した. 今回のエラー分析では, 他のタスクとの関連性に関する考察が不足している. 形態素解析, 文節の係り受け解析, 述語項構造解析などの前処理の結果が WSD にどのように影響するかを分析することが今後の課題である. WSD の誤りが機械翻訳や情報検索などの上位のタスクにどの程度影響を与えるのかも検証するべきであろう. 例えば, 機械翻訳では, 訳語を語義とみなせば翻訳と同時に語義の曖昧性も解消していることになるが, Cai et al. は統計的機械翻訳による訳語選択の正解率と WSD のためだけに機械学習されたシステムの正解率を比較し, 後者の方が高いことから, WSD のモジュールを統計的機械翻訳に組み込むことは有望であると主張している [4]. このように WSD の技術が様々な自然言語処理応用システムにどの程度貢献するかを調査することは重要な課題である. また, 4.1.5 で述べた消去法による WSD を実現する方法も検討したい.

参考文献

- [1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [2] Jordan Boyd-Graber, David Blei, and Xiaojin Zhu. A topic model for word sense disambiguation. In *Proceedings of the the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1024–1033, 2007.
- [3] Jun Fu Cai, Wee Sun Lee, and Yee Whye Teh. Improving word sense disambiguation using topic features. In *Proceedings of the the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1015–1023, 2007.
- [4] Marine Carpuat and Dekai Wu. Evaluating the word sense disambiguation performance of statistical machine translation. In *Proceedings of the the Second International Joint Conference on Natural Language Processing*, pp. 120–125, 2005.
- [5] Moldovan Dan and Adrian Novischi. Word sense disambiguation of WordNet glosses. *Computer Speech and Language*, Vol. 18, No. 3, pp. 301–317, 2004.
- [6] 藤井文明, 新納浩幸, 佐々木稔. 語義識別の誤り原因の調査とオンザフライの類似語判定. 言語処理学会第 10 回年次大会, pp. 753–756, 2004.
- [7] Sanae Fujita and Akinori Fujino. Word sense disambiguation by combining labeled data expansion and semi-supervised learning method. *Transactions on Asian Language Inforamtion Processing, Association for Computinng Machinery (ACM)*, Vol. 22, No. 7, pp. 676–685, 2013.
- [8] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系 CD-ROM 版. 岩波書店, 1999. resource.
- [9] 西尾実, 岩淵悦太郎, 水谷静夫. 岩波国語辞典 第五版. 岩波書店, 1994. resource.
- [10] Manabu Okumura, Kiyooki Shirai, Kanako Komiya, and Hikaru Yokono. On Semeval-2010 Japanese WSD task. *自然言語処理*, Vol. 18, No. 3, pp. 293–307, 2011.
- [11] 新納浩幸, 白井清昭, 村田真樹, 福本文代, 藤田早苗, 佐々木稔, 古宮嘉那子, 乾孝司. 語義曖昧性解消の誤り分析. 言語処理学会第 21 回年次大会ワークショップ自然言語処理におけるエラー分析, 2015.
- [12] 白井清昭. SENSEVAL-2 日本語辞書タスク. *自然言語処理*, Vol. 10, No. 3, pp. 3–24, 2003.
- [13] Kiyooki Shirai and Takayuki Tamagaki. Word sense disambiguation using heterogeneous language resources. In *First International Joint*

Conference on Natural Language Processing, pp. 614–619, 2004.

- [14] 玉垣隆幸, 白井清昭. 読解支援システムのための語義曖昧性解消に関する研究. 言語処理学会第9回年次大会, pp. 481–484, 2003.
- [15] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 189–196, 1995.
- [16] 国立国語研究所コーパス開発センター. 『現代日本語書き言葉均衡コーパス』利用の手引 (第 1.0 版), 2011.