ラベルなしデータの二段階分類とアンサンブル学習に基づく 半教師あり日本語語義曖昧性解消

井上 裁都[†]・斎藤 博昭[†]

本稿では、パラメータ調整を簡略化したブートストラッピング的手法による日本語語義曖昧性解消を提案する。本稿で取り上げるブートストラッピングとは、ラベルなしデータを既存の教師あり学習手法を用いて分類し、その中で信頼度の高いデータをラベル付きデータに加え、この手順を反復することによって分類の性能を向上させる半教師あり学習手法である。従来のブートストラッピングによる語義曖昧性解消においては、プールサイズ、ラベル付きデータに追加するラベルなしデータの事例数、手順の反復回数といったパラメータをタスクに合わせ調整する必要があった。本稿にて提案する手法はヒューリスティックと教師あり学習（最大エントロピー法）によるラベルなしデータの二段階の分類、および学習に用いるラベルなしデータの条件を変えた複数の分類器のアンサンブルに基づく。これにより必要なパラメータ数は一つになり、かつパラメータの変化に対し頑健な語義曖昧性解消を実現する。SemEval-2 日本語タスクのデータセットを用いたベースラインの教師あり手法との比較実験の結果、パラメータの変化に対し最高で 1.8 ポイント、最低でも 1.56 ポイントの向上が見られ、提案手法の有効性を示せた。

キーワード：語義曖昧性解消，半教師あり学習，ブートストラッピング，最大エントロピー法，アンサンブル学習，頑健性

Semi-Supervised Japanese Word Sense Disambiguation Based on Two-Stage Classification of Unlabeled Data and Ensemble Learning

TATSUKUNI INOUE[†] and HIROAKI SAITO[†]

In this paper, we propose a bootstrapping-like method which eases optimal and empirical parameter selection for Japanese word sense disambiguation. Bootstrapping means, in this paper, semi-supervised learning methods based on the following procedures: (1) train a classifier on labeled examples, (2) use the classifier to select confident unlabeled examples, (3) add them to the labeled examples, (4) repeat steps 1–3. Traditional bootstrapping methods require empirical selection for the parameters including the pool size, the number of the most confident examples and the number of iterations. Our method uses two-stage unlabeled example classification based on heuristics and a supervised method (Maximum Entropy classifier) and combines a series of classifiers along a sequence of varying conditions. This method requires only

[†] 慶應義塾大学大学院理工学研究科, Graduate School of Science and Technology, Keio University

one parameter and enables parameter robust word sense disambiguation. Experiments compared with the baseline supervised method on the Japanese WSD task of SemEval-2 shows that our method obtained accuracy improvement between 1.8 and 1.56 points.

Key Words: *Word sense disambiguation, Semi-supervised learning, Bootstrapping, Maximum entropy, Ensemble learning, Robustness*

1 はじめに

語義曖昧性解消は古典的な自然言語処理の課題の一つであり、先行研究の多くは教師あり学習により成果を挙げてきた (Màrquez, Escudero, Martínez, and Rigau 2004; Navigli 2009). しかし、教師あり学習による語義曖昧性解消においてはデータスパースネスが大きな問題となる. 多義語の語義がその共起語より定まるという仮定に基づけば、一つの多義語と共起し得る単語の種類が数万を超えることは珍しくなく、この数万種類のパターンに対応するために十分な語義ラベル付きデータを人手で確保し、教師あり手法を適用するのは現実的でない. 一方で語義ラベルが付与されていない、いわゆるラベルなしのデータを大量に用意することは、ウェブの発展、学術研究用のコーパスの整備などにより比較的容易である. このような背景から、訓練データと大量のラベルなしデータを併用してクラス分類精度を向上させる半教師あり学習、または訓練データを必要としない教師なし学習による効果的な語義曖昧性解消手法の確立は重要であると言える.

本稿では半教師あり手法の一つであるブートストラッピング法を取り上げ、従来のブートストラッピング法による語義曖昧性解消手法の欠点に対処した手法を提案する. ブートストラッピング法による語義曖昧性解消においては主に Self-training (自己訓練) (Nigam and Ghani 2000) と Co-training (共訓練) (Blum and Mitchell 1998) の二つのアプローチがある (Navigli 2009). まずこれらの手法に共通する手順を述べると次のようになる.

一般的ブートストラッピング手順

Step 1 ラベルなしデータ U から事例 P 個をランダムに取り出し U' を作る.

Step 2 ラベル付きデータ L を用いて一つまたは二つの分類器に学習させ U' の事例进行分类する.

Step 3 Step 2 で分類した事例より分類器毎に信頼性の高いものから順に G 個を選び、 L に加える.

Step 4 Step 1 から R 回繰り返す.

Self-training と Co-training の違いは、前者は Step 2 で用いる分類器は一つであるのに対し、後者は二つ用いる点にある. また Co-training においては二つの独立した素性集合を設定し、各

分類器を一方の素性集合のみを用いて作成する. Co-training においてこのように設定するのは, Step 3 において追加する事例を一方の素性のみから決定することから, 追加事例のもう一方の素性を見たとき新しい規則の獲得が期待できるためである. Self-training と Co-training の欠点はいずれも性能に影響するパラメータが多数存在し, かつこれらのパラメータを最適化する手段がないことである. 具体的には Step 1 のプールサイズ P , Step 3 の L に加える事例の個数 G , 手順の反復回数 R は全てパラメータであり, タスクに合わせた調整を必要とする.

本稿では, ラベル付きデータとラベルなしデータを同時に活用しつつも, パラメータ設定をほとんど不要とする新しい手法を提案する. 本手法はまずヒューリスティックと教師あり学習で構築した分類器によるラベルなしデータの二段階の「分類」を行う. ここで「分類」とは語義曖昧性解消を行い, 語義ラベルを付与することを意味する. 本稿では以後特に断りがない限り, 分類とはこの語義ラベル付与のことを指す. 二段階分類したラベルなしデータの中で条件を満たすデータはオリジナルのラベル付きデータに加えられる. その結果, パラメータ設定がほぼ不要なブートストラッピング的半教師あり手法による語義曖昧性解消を実現する. さらに追加するラベルなしデータの条件を変えることで複数の分類器を作成し, アンサンブル学習することで, パラメータの変化に頑健な分類器を生成する.

本稿の構成は以下の通りである. 2 節にて関連研究および本研究の位置付けを述べる. 3 節にて提案手法およびその原理を並行して述べる. 4 節にて SemEval-2 日本語タスク (Okumura, Shirai, Komiya, and Yokono 2010) のデータセットに提案手法を適用した実験の結果を示す. 5 節にて結論を述べる.

2 関連研究

本節ではまず 1 節でブートストラッピング手法として挙げた Self-training および Co-training を用いた語義曖昧性解消の先行研究を概観する. また, アンサンブル学習に基づく語義曖昧性解消には教師あり学習のアンサンブル, 教師なし学習のアンサンブル, 半教師あり学習のアンサンブルに基づいた手法が提案されており, これら先行研究を併せて概説する.

2.1 ブートストラッピングに基づく研究

Self-training に基づいた語義曖昧性解消の先駆けとしては Yarowsky の研究 (Yarowsky 1995) が挙げられる. Yarowsky は「語義はその語の連語より定まる (one sense per collocation)」「語義はその語を含む談話より定まる (one sense per discourse)」という二つのヒューリスティックに基づき, ラベルなしデータに反復的にラベル付けするアルゴリズムを提案した. この手法は二つの観点からラベル付けをするため, Co-training の一種であるとも見ることもできる. また, このヒューリスティックに基づいた Yarowsky のアルゴリズムは Abney (Abney 2004) により,

目的関数の最適化問題として定式化されている。

Co-training を用いた語義曖昧性解消の早期の例としては新納の報告 (新納 2001) がある。新納は Co-training を適用するにあたり、二組の素性集合の独立性を高めるため、ラベル付きデータに追加するラベルなしデータを素性間の共起性に基づいて選択する手法を提案した。結果、日本語の語義曖昧性解消において通常の Co-training よりも性能が向上したと報告した。

Mihalcea は Co-training と Self-training の両方を語義曖昧性解消に適用し、1 節にて述べたパラメータの影響について調査した (Mihalcea 2004)。この報告ではパラメータの自動での最適化はできず、最適な設定と自動による設定に大きな差があったと報告している。また、Mihalcea は同じ報告の中でスムージングした Co-training および Self-training を提案した。これは手順の反復のたびに生成される各分類器の多数決より語義判定しブートストラッピングするという方式であり、ブートストラッピングとアンサンブル学習の組合せの一種と見ることができる。この方式は通常のブートストラッピングよりも性能が向上したと報告された。

以上の手法は 1 節で述べたようなパラメータをタスク (データセット) に合わせ調整しなければならないという大きな課題がある。

Niu らは Zhu らの提案したラベル伝播手法 (Zhu and Ghahramani 2002) に基づいた半教師あり手法による語義曖昧性解消について調査した (Niu, Ji, and Tan 2005)。ラベル伝播は事例を節点とする連結グラフを考え、重み付きの辺を通してラベルありの事例からラベルなしの事例へとラベル情報を反復的に伝播させる。そして伝播の収束結果よりラベルを推定する。この手法は Senseval-3 (Mihalcea and Edmonds 2004) English lexical sample タスクのデータセットに適用した結果、従来の教師あり学習と比較して著しい成果は得られなかったとしている。

Pham らは Co-training と Smoothed Co-training (Mihalcea 2004) に加え、Spectral Graph Transduction (SGT) (Joachims 2003) および SGT と Co-training を組合せた Co-training の語義曖昧性解消への適用を調査した (Pham, Ng, and Lee 2005)。Transduction とは訓練データから分類器を生成せず、直接テストデータにラベル付けする推論方法である (Vapnik 1998)。SGT は k 近傍法の Transductive 版であるとされる。SGT は k 近傍法の応用であるため、 k がパラメータとなり、かつ k は性能に与える影響が大きいと報告されている。よって Pham らの調査した手法全てにはパラメータ設定の問題が存在していることになる。

手法のアンサンブルを含まないブートストラッピングによる語義曖昧性解消の研究の最後として小町らの報告 (小町, 工藤, 新保, 松本 2010) を挙げる。小町らはブートストラッピング手法の一つである Espresso (Pantel and Pennacchiotti 2006) に対しグラフ理論に基づいて意味ドリフト (Curran, Murphy, and Scholz 2007) の解析を行った。意味ドリフトは、語義曖昧性解消の観点から考えると、どのような語義の語も持つ素性をジェネリックパターンと考え、ジェネリックパターンを持つ (信頼性が低いとされるべき) ラベルなしデータに対し手順の反復過程においてラベルが与えられることにより、反復終了後に生成される分類器の性能が低下してし

もう現象と解釈できる。この問題への対処のため小町らは二つのリンク解析的関連度算出法を適用した。この手法は意味ドリフトに頑健かつパラメータ数が一つでさらにその調整が比較的容易という利点を持つ。Senseval-3 English lexical sample タスクのデータセットに手法を適用した実験の結果、小町らの手法は類似したグラフ理論的手法である HyperLex (Véronis 2004) や PageRank (Agirre, Martínez, de Lacalle, and Soroa 2006) と比較して高い性能が得られたと報告している。

2.2 アンサンブル学習に基づく研究

教師あり学習のアンサンブルに基づく研究としては、AdaBoost を用いた (Escudero, Màrquez, and Rigau 2000), 素性として用いる文脈の大きさを変えて複数の Naive Bayes 分類器をアンサンブルした (Pedersen 2000), 六種の分類器の組合せによる (Florian, Cucerzan, Schafer, and Yarowsky 2002), 二段階の分類器の出力の選択に基づいた (Klein, Toutanova, Ilhan, Kamvar, and Manning 2002), 複数の Naive Bayes 分類器の出力の比較に基づいた (Wang and Matsumoto 2004) が挙げられる。ここでは Wang らの手法をより詳しくみる。Wang らはまず Pederson と同様に素性として用いる文脈の大きさ、つまり目標の多義語前後 k 語以内の語を素性として用いるとして、 k を変えることで複数の Naive Bayes 分類器を作成する。次にラベル付きデータを各分類器にて分類する。各要素がこの各分類器による分類結果であるベクトルを decision trajectory と呼ぶ。最後に各ラベル付きデータから得た decision trajectory の集合を訓練データとし、これらと入力から得た decision trajectory の類似度に基づいて入力の語義を判定する。Wang らの手法は中国語の語義曖昧性解消実験の結果、Pederson の手法などと比較して最も良い結果が得られたと報告した。

教師なし手法のアンサンブルの例としては Brody らの研究 (Brody, Navigli, and Lapata 2006) が挙げられる。Brody らは過去に語義曖昧性解消において有効と報告された教師なし手法である Extended Gloss Overlap (Banerjee and Pedersen 2003), Distributional and WordNet Similarity (McCarthy, Koeling, Weeds, and Carroll 2004), Lexical Chains (Galley and McKeown 2003) および Structural Semantic Interconnections (Navigli and Velardi 2005) を組合せた手法を提案した。この手法は組合せに用いた各手法と比較し、より良い結果が得られたと報告されている。

最後に本稿で提案する手法に最も関連の深いブートストラッピング手法のアンサンブルを行った Le らの研究 (Le, Shimazu, Huynh, and Nguyen 2008) について述べる。Le らは我々と同様に従来のブートストラッピングによる語義曖昧性解消の問題点に対する解決法を提案した。Le らが解決法を提案した問題点は、(1) ラベル付きデータのラベル毎のデータ数の偏り、(2) ラベル付きデータに追加するラベルなしデータ決定の基準、(3) 手順の反復の停止条件および最終分類器の作成法の三つである。ここで問題 (2) は 1 節にて述べたパラメータ G の決定法、問題 (3) はパラメータ R の決定法とも換言できる。Le らはこれらの解決のため、追加するラベルなし

データのリサイズ, 複数のデータ追加条件の閾値の設定および対応する複数の追加データ集合の設定, 訓練データを用いた追加データの評価および手順反復停止条件の設定, そして追加データと教師あり学習手法別の各分類器のいくつかの組合せ法を提案した. ここで追加データの評価と手順反復停止条件の設定の手法は, Zhou らが提案した Tri-training 法 (Zhou and Li 2005) で用いられた手法を参考に設定している. Tri-training は Co-training を発展させた手法であり, Co-training と異なりパラメータ設定を不要とする特徴がある. 実験は Senseval-2 (Edmonds and Cotton 2001) および Senseval-3 の English lexical sample タスクのデータセットを用いて行われ, 従来の教師あり手法と比較し最大で 1.51 ポイントの精度向上が見られたと Le らは報告した.

2.3 本研究の位置付け

2.1 節と 2.2 節を踏まえた上での本研究の位置付けは以下の通りである. まず, 小町らの手法はパラメータ設定が容易という利点があるが, 他の教師あり手法と組合せるのが困難なのが問題点である. 高性能な教師あり手法を用いず, さらに性能を向上させるのは難しい. また, Le らの手法の難点として手順の反復停止条件の設定が挙げられる. これは, 教師あり学習を用い追加データを訓練データとして分類器を作成しオリジナルのラベル付きデータを分類して得られるエラー率, および追加データの総数に基づき設定される. 具体的には次の式を用い追加データの評価値 q を求める.

$$q = m(1 - 2\eta)^2 \quad (1)$$

ここで, m は追加データの総数, η はエラー率を示す. この q が前回の値よりも小さければ反復は停止する. しかし反復の停止にこの条件が用いられる具体的根拠は示されていない. 単に反復を自動的に停止するためと述べられているだけである. このためこの停止条件が最適であるかどうか疑問が残る.

そこで本研究の立場だが, まず本稿ではこの停止条件の追究はしない方針とする. しかし, 1 節にて目標としたようにブートストラッピングにおけるパラメータの削減は達成を目指す. そこで本研究では手順の反復回数 (パラメータ R) を一回に留めるという方針を採用. この方針には次の利点が考えられる.

- 手順の回数が固定され計算時間の予測が立てやすい.
- ラベル付きデータへのラベルなしデータの追加が一度のみとなるため, 追加されたデータに対し分析, 考察を加えやすい.
- 反復 1 回目の精度を向上させることで, 手法を複数反復できるように拡張したとき更なる精度向上が見込める.

以上の検討に基づき, 本研究では反復を伴わず, かつ教師あり学習手法を併用した高性能なブートストラッピング的手法を確立する. また反復回数を一回にすることは, 反復回数以外の

パラメータを削減することにもつながる。詳しくは3節にて述べる

3 提案手法

提案手法は1節で述べたラベルなしデータの二段階の分類とその結果を用いたアンサンブル学習による最終分類器作成の全三段階からなる。手法の流れを次に、また本手法に基づくシステム全体像を図1に示す。

Stage 1 ヒューリスティックによる分類

(手掛り語獲得, 手掛り付き事例抽出・分類1回目)

Stage 2 教師あり学習手法による分類

(手掛り付き事例分類2回目, 二次分類器作成)

Stage 3 アンサンブル学習による最終分類器作成

まず図1の見方を述べる。長方形の囲みは各種処理, 楕円の囲みは各処理の出力を示す。各種処理の中には語義分類の処理が三回登場するが, これらの括弧内の分類器はそれぞれの分類処理のために作成し使用する分類器を示している。実線の矢印は各処理において入出力されるデータの流れを示す。点線は入出力するデータへの処理に利用するデータであり, 分類器作成のために訓練データとして用いるデータも含まれる。図1における訓練データとはオリジナルのラベル付きデータを指す。本システムにおける最終的な語義曖昧性解消の対象は, テストデータとして図1のように入力し, その結果は語義分類最終結果として出力される。

図1に基づく本システムの概観は次のようになる。まず本システムの最初の手順は手掛り語の獲得である。手掛り語は訓練データから抽出する形で獲得する。第二の手順はラベルなしデー

図1 提案手法システム全体像

タからの手掛り付き事例の獲得である。手掛り付き事例の獲得は、手掛り語を用いたラベルなしデータからの手掛り付き事例の抽出・分類（1回目）、訓練データを用いて作成された一次分類器による手掛り付き事例の語義分類、一次分類器による分類結果に基づく手掛り付き事例の分類（2回目）といった多段の手順により実現される。最後の手順はテストデータの語義曖昧性解消である。これは獲得した手掛り付き事例と訓練データを用いた二次分類器の作成、および異なる条件で作成された複数の二次分類器によるテストデータの分類結果に基づく最終分類器の判定より構成される。

本節では、以後本手法を全三段階に区切り、詳述していく。ここで一つ注意点がある。それは今、本システムを上述のように手掛り語、手掛り付き事例、語義分類最終結果と出力されるデータに着目し、手順を三つに区切ったが、これは以後に述べる三段階の手順と対応関係にないということである。具体的には、手法第一段階は図1の手掛り語獲得から手掛り付き事例抽出・分類1回目まで、手法第二段階は一次分類器による手掛り付き事例の分類から二次分類器によるテストデータの語義分類まで、手法第三段階は最終分類器によるテストデータの語義分類のみと対応する。

手法第一段階はヒューリスティックによるラベルなしデータの分類として一括し、手掛り語ならびに手掛り付き事例の詳細と併せて3.1節にて詳述する。手法第二段階はラベルなしデータから抽出した手掛り付き事例の教師あり学習手法による分類とその結果に基づく二次分類器の作成として括り出し、詳細を3.2節にて述べる。手法第三段階はアンサンブル学習による最終分類器の作成として括り、3.3節にて詳述する。3.4節では本手法のまとめをする。このような手法の全三段階の区切りは、4節にて述べる実験結果の考察において意味を持つことになる。

なお本稿では、訓練データ、テストデータおよびラベルなしデータはいずれも UniDic¹を用いて形態素解析済み²であり、訓練データにおいてラベルは各形態素に付与されているものとする。また本稿では便宜上、形態素を単語または語とも呼ぶことにする。

3.1 Stage 1：ヒューリスティックによる分類

分類第一段階は訓練データからの手掛り語の獲得ならびにヒューリスティックによるラベルなしデータからの手掛り付き事例の抽出・分類（1回目）からなる。

3.1.1 手掛り語の獲得

本手法で獲得する手掛り語 W_{ts} とは、訓練データ L において語義ラベルが付与された対象語³ t の前後 n_w 語以内において共起する内容語⁴ の表層形であり、かつ与えられた訓練データ

¹ <http://www.tokuteicorpus.jp/dist/>

² SemEval-2 日本語タスクのデータセットは UniDic を用いた自動解析およびその人手での修正が施されている。

³ 本稿では語義曖昧性解消の対象語を単に「対象語」と呼ぶ。

⁴ 具体的には4節参照のこと。

内で共起する t に付与された語義ラベルが必ずある一つの語義ラベル s に定まる語の集合とする。後者の条件は, s が付与された t を t_s とすると, 形態素 w_j 共起の下で t が t_s である確率を $p(t_s|w_j)$ とすると,

$$p(t_s|w_j) = 1 \quad (2)$$

を満たす w_j であることも書き換えることができる。前者の条件にある窓幅 n_w については 3.3 節で詳述する。

このような条件を満たす $w_{ts} \in W_{ts}$ が語義ラベルが付与されていない t と共起したとき, この t は単純に式 (2) より t_s である可能性が高いと考え, 以後 W_{ts} は t の語義曖昧性解消の手掛りとして利用する。また表層形を条件としたのは, 基本形, 品詞といった情報は表層形と比べ情報の粒度が荒く, 表層形の方が手掛りとしての信頼性が高いと考えたことによる。

もし式 (2) を満たす w_j が L 全体で一度だけ出現する語である場合, t が t_s に決定付けられる可能性は低いとも考えられるが, これは二度以上出現する語の場合も大差はないと筆者は考える。 $p(t_s|w_j)$ は語義曖昧性解消において単純ベイズや決定リストのルールの信頼度などにしばしば用いられ⁵, その中で $p(t_s|w_j)$ をスムージングして用いる例もいくつかある (Yarowsky 1995; 八木, 野呂, 白井, 徳永, 田中 2001; 鶴岡, 近山 2002)。しかしこの場合は最適な閾値を求める必要があり, 問題がかえって難しくなってしまう。このため, 今回は単純な式 (2) を w_{ts} の条件とした。

なお上述の w_{ts} は添え字が示すように共起する t_s に付与されていた s の情報を含む。よって実際の手掛り語獲得では, 例えば対象語が「相手」, $n_w = 2$, 訓練データの一つに「相手に取って不足はない」があり, この文中の「相手」に “117-0-0-3” という語義 ID⁶ が付与されていたとする。このとき「取っ」という語が訓練データにおいて “117-0-0-3” の語義の「相手」とのみ共起するのであれば, この訓練データからは《取っ, 117-0-0-3》という二つ組一つを抽出する。

3.1.2 手掛り付き事例抽出・分類 (1 回目)

ここでは, まず手掛り付き事例抽出の手順を述べる前に SemEval-2 日本語タスクにおける対象語 t の表記ゆれへの対処について述べる。SemEval-2 においては t について与えられる情報は訓練データ L を除くと与えられた辞書に記述された見出し語 H_t と語義の語釈文 D_t のみであり, t の表記に関する情報は充分には与えられない。例えば, t の一つに「子供」があるが, これは他にも「子ども」「こども」といった表記があるのに対し, H_t には「子ども」という表

⁵ ここでは w_j は任意の素性である。

⁶ ここに示した ID は SemEval-2 日本語タスクにて用いられたもの。上位二つの数字は見出し語の ID, 残り二つはそれぞれ語義の大分類, 中分類の ID を示す。なお “117-0-0-3” の辞書定義文は「自分と対抗して物事を争う人」である。

記の情報は与えられていない。この問題に対処しない場合、ラベルなしデータから t の事例を十分な数獲得できないだけでなく、 t の表記により語義の傾向が変わる場合も考えられ、抽出する事例の語義に偏りが生まれる可能性も考えられる。このため、UniDic の辞書を用いて、以下の手順で t の取り得る表記（表層形） E_t の獲得を行った。なお下記の Step 2 では実際には E_{t0} からひらがなのみで構成される表層形を除外して V_t を抽出している。これはこのような語は語彙素の同定が困難であり表層形獲得精度の低下を招くためである。また $e_t \in E_t$ は品詞細分類の情報も合わせて獲得し、以後 t の事例としてこの二つ組の情報が一致するものを獲得する。

対象語表層形の獲得手順

Step 1 t に対し L と H_t から獲得可能な全ての表層形 E_{t0} を抽出する。

Step 2 E_{t0} と対応する語彙素 V_t を UniDic の辞書から抽出する。

Step 3 V_t の全表層形 E_{t1} を UniDic の辞書から抽出する。

Step 4 E_{t0} および E_{t1} を合わせて対象語の表層形 E_t として獲得する。

続いて、獲得した手掛り語 W_{ts} および対象語の表層形 E_t を用いてラベルなしデータ U より手掛り付き事例の抽出および 1 回目の分類を行う。手順を以下に示す。

手掛り付き事例抽出・分類（1 回目）の手順

Step 1 U から $e_t \in E_t$ と一致する表層形の形態素を探索し、発見したら i_{t0} とする。

Step 2 i_{t0} の前後 n'_w 語以内に $w_{ts} \in W_{ts}$ が共起する場合、 i_{t0} を手掛り付き事例 $i_{ts1} \in I_{ts1}$ として抽出する。

手掛り付き事例 I_{ts1} とは対象語 t の前後に手掛り語 w_{ts} が共起する事例を指し、上記手順より抽出される。また、ここで抽出される I_{ts1} の s は w_{ts} の添え字の s であり、 i_{t0} に s を付与することで分類したとみなすことができる。よって上記手順では、手掛り付き事例の抽出と分類を同時に行っていると解釈できる。一方で、 i_{t0} の集合 I_{t0} と $\bigcup_s I_{ts1}$ の差集合は語義ラベルが付与されないということで、語義判定不可に分類されたと考えることもできる。

上記手順を具体例を挙げ説明すると次のようになる。 E_t に「相手」、 W_{ts} に《取っ, 117-0-0-3》が含まれているとし、 U より「ゼネコン三十一社を相手取って一人あたり三千三百万円の損害賠償を求めた」という文に対し上記手順を適用するとする。また、 $n'_w = 2$ とする。この場合、文中の「相手」が i_{t0} となり、 t の前後 2 語以内に「取っ」が共起するため、 $s = \text{“117-0-0-3”}$ とし、 i_{t0} を手掛り付き事例 i_{ts1} として抽出する。

さて、ここでパラメータ n'_w についてであるが、これは W_{ts} 獲得に用いるパラメータ n_w とは区別する。さらに n'_w の値は上述の例と同様に“2”と固定する。この 2 という数は 3.2 節で述べる教師あり学習による分類の素性として対象語前後“2”語以内の形態素を用いることと対応するのだが、その理由を列挙すると以下のようになる。

- (1) W_{ts} 獲得に用いる訓練データ L は本タスクにおいて数少ない信頼できるデータである。したがって、 L からは出来る限り多くの特徴を抽出したい。
- (2) 一方、ラベルなしデータ U は多量に存在するが、これを自動的に分類したデータは当然ながら必ずしも信頼できるわけではない。
- (3) 反復回数一回でなるべく信頼性が高くかつ十分な数のデータの獲得が望ましい。
- (4) n'_w を教師あり学習の素性抽出の範囲と一致させた場合、抽出した手掛り付き事例 i_{ts1} の素性に必ず w_{ts} が含まれる。このため、 i_{ts1} を教師あり手法で再分類したとき高精度の分類が期待できる。
- (5) w_{ts} は n_w に関わらず式 (2) を満たす。つまりある程度の範囲までは n_w を大きくすることで信頼性を維持しつつ多数の w_{ts} を獲得できる。
- (6) w_{ts} が十分な数あれば、一度の処理で多数の U を分類しやはり十分な数のデータを L に加えることができる。

上述の理由には従来法の欠点と提案法の利点の両方が含まれている。その対応関係は、理由 (4) は (2) への対処であり、(5) は (4) の補足かつ (1) への対処であり、(6) は (5) を踏まえた (3) への対処となる。またここに述べた理由は、3.2 節にて述べる手掛り付き事例分類 2 回目において、1 節で述べたパラメータ P, G, R が削減可能となる理由にもなる。詳しくは 3.2 節にて改めて述べる。

以上のアルゴリズムをもって、本手法の第一段階とする。節題の通り、本処理は経験則に基づく部分が多い。しかし、本処理は以降の処理においても必要とされる性質を備えている。これらは 3.2 節および 3.3 節にて詳述する。

3.2 Stage 2：教師あり学習手法による分類

分類第二段階では 3.1 節で抽出・分類した手掛り付き事例に対し、オリジナルの訓練データから得られる一次分類器を用いて 2 回目の分類を行う。そして、その結果得られる手掛り付き事例を用いて二次分類器を作成する。

3.2.1 教師あり学習手法

本手法で用いる教師あり学習手法は最適化に L-BFGS (Liu and Nocedal 1989) を用いた最大エントロピー法 (Nigam, Laerty, and McCallum 1999) とした。この理由は SemEval-2 日本語タスクフォーマルラン参加チームの一つの報告 (Fujita, Duh, Taira, and Shindo 2010) に最大エントロピー法が有効というものがあったためである。また学習の素性も Fujita らの報告を参考に次のように設定した。

- 範囲
対象語前後 2 語以内

- 1 グラム素性
形態素の表層形
形態素の基本形⁷
- 2 グラム・3 グラム・対象語を含むスキップ2 グラム素性
形態素の表層形
形態素の基本形
形態素の品詞と対象語との相対位置の組合せ
形態素の品詞細分類と対象語との相対位置の組合せ

具体的には、対象語「相手」の事例「相手に取って不足はない」に対しては次の素性が獲得できる。下記例中の“*”を含む素性はスキップ2 グラムを示す。また品詞に付与されている番号は対象語との相対位置である。

相手, に, 取っ, トル, 相手 に, に 取っ, 相手 に 取っ, 相手 * 取っ, に トル, 相手 に
トル, 相手 * トル, 名詞₀ 助詞₁, 助詞₁ 動詞₂, 名詞₀ 助詞₁ 動詞₂, 名詞₀ * 動詞₂,
名詞-普通名詞-一般₀ 助詞-格助詞₁, 助詞-格助詞₁ 動詞-一般₂, 名詞-普通名詞-一般₀ 助
詞-格助詞₁ 動詞-一般₂, 名詞-普通名詞-一般₀ * 動詞-一般₂

3.2.2 手掛け付き事例分類 (2 回目)

前述の学習手法, 素性, そして訓練データ L を用いて一次分類器 C_1 を作成し, C_1 を用いて 3.1 節で抽出・分類した手掛け付き事例 I_{ts1} を再分類する。この分類 2 回目の結果が分類 1 回目の結果と一致する, つまり C_1 の $i_{ts1} \in I_{ts1}$ の分類結果を $c_1(i_{ts1})$ としたとき, $c_1(i_{ts1}) = s$ である場合, i_{ts1} を $i_{ts2} \in I_{ts2}$ とし L に加え, これを用いて C_1 同様に二次分類器 C_2 を作成する。

3.1 節で述べたように i_{ts1} はその素性に必ず手掛け語 w_{ts} を含む。そのため $c_1(i_{ts1})$ は s と一致する可能性が高いが, 実際に一致を確認し, 一致しなければ C_2 作成においてこの手掛け付き事例は使わない。この結果, C_2 作成に用いられる I_{ts2} のラベル s は信頼性の高いものとなる。

このシステムの重要な点は単に二種類の分類手法の結果が合致するものを選択することではない。そうだとすれば, 分類 1 回目は一般的な教師あり学習手法を用いても良いことになってしまう。重要なのは分類 2 回目にて C_1 が高い精度で分類可能な事例 I_{ts1} を分類 1 回目において選択していることにある。つまり, 分類 1 回目が分類 2 回目の精度向上を明確に支援していることがポイントである。これにより I_{ts2} 全てを L に加えても信頼性は保持され, 同時に十分な数のブートストラッピングが可能になる。これは従来法において必要だった 1 節に挙げたパラメータ, ラベル付きデータ L に加える事例の個数 G および手順の反復回数 R を決めることなしに適切な事例を L に加えられることも意味する。なぜなら, G を定めずとも全事例を L に加えればよく, R を定めずとも一度の実行で十分な数の事例の獲得が可能だからである。また

⁷ 本研究では用言の基本形のカナ表記を「基本形」とする。詳しくは 4 節参照のこと。

プールサイズ P については, Blum らの考察 (Blum and Mitchell 1998) から考えるとブートストラッピングの反復において意味を持つ値であると思われる. よって, 反復回数 1 の本手法は単にプールを設定する必要がなく, 事例を全てのラベルなしデータ U から抽出することで処理できる.

また, W_{ts} は $n_w > 2$ であれば L の素性にはない語 (つまり対象語から 3 語以上離れた位置にある語) を含むことから, I_{ts1} の素性はやはり L の素性にはない語を含む. すると I_{ts1} が分類 2 回目の結果, L に追加されれば, 上述の通り分類 2 回目の信頼性は高いと言えるため, C_2 は C_1 と比べ正しく分類できる U が増える可能性が高い. したがって, 本手法は U の L への追加における効率性が高い手法であるとも言えることができる.

上述の性質は Co-training との類似性を指摘することもできる. Co-training は 1 節で述べたように二つの素性集合のうち一方のみに基づいて分類することで他方の素性について新しい規則の獲得が期待できるのが特徴である. 本手法では, W_{ts} と C_1 に用いる素性の二種類の素性を実質的に両方考慮して手掛け付き事例を分類している. しかし, C_2 に用いる素性は後者の素性のみである. つまり, 一方は他方の一部ということになるが, 一方の素性で分類した結果を他方の素性を用いる分類器の訓練データに加えるという点では Co-training と共通する. その一方で, 提案手法は L に追加する U の分類に二種類, つまり全ての素性を用いており, 一方のみを使う場合と比較すると分類結果の信頼性が高いという利点がある. このような変則的な Co-training と通常の Co-training の間にどのような差異が生まれるかは未調査だが, ここに述べた性質は性能の向上に結び付くと期待される.

本処理の直感的な意味としては, C_1 にまず簡単な問題を解かせ, その結果を C_2 の学習に利用していると解釈できる. 一方で従来のブートストラッピングは, 難易度がランダムな問題を複数解かせ, その中でシステムが自信を持って答えられるものから学習すると考えられる. しかし後者の場合, 回答に対し「間違った自信」を持ってしまい, 結果として不適切な学習をしてしまう危険性があり得る. 前者の, つまり提案した手法は, 確実ではないが C_1 が解くのが簡単であろう問題を選択しており, この危険性はいくらか低減していると推測される. この推測が正しいとすれば, C_1 に提示する問題を選択する分類 1 回目の処理は重要な意味を持つことになる. また, ここで提示するのは勿論ラベルなしの文章であるが, 見方を変えると「良い文章」をシステムに提示することでより良い学習が可能になると考えられ, 興味深い.

3.3 Stage 3: アンサンブル学習による最終分類器作成

本節では 3.2 節で作成した二次分類器 C_2 をアンサンブルして最終分類器 C_3 を作成する方法を述べる. アンサンブルには 3.1 節の手掛け語抽出において決定法を保留していた窓幅 n_w を利用する. すなわち, n_w をパラメータとする二次分類器を $C_2(n_w)$ とし, n_w を変化させた $C_2(n_w)$ を複数組合せ C_3 とする. 組合せの方法は各最大エントロピー分類器が出力する各ラベルの推

定確率の中で最高値を出力した分類器の判定を採用する方式とする．つまり入力を $\mathbf{x}$ ， $C_2(n_w)$ が $\mathbf{x}$ に対し出力する語義 s である推定確率を $p(s|\mathbf{x}, n_w)$ とすると，

$$s_*(\mathbf{x}) = \arg \max_s \left[\max_{n_w} p(s|\mathbf{x}, n_w) \right] \quad (3)$$

より求まる $s_*(\mathbf{x})$ を C_3 の出力とする．

式 (3) の方式で良い結果が得られる根拠は手掛り語 W_{ts} の条件の一つである式 (2) にある．式 (2) の制約の下で n_w の値を大きくしたとき，得られる W_{ts} に以下の二つの変化が見られる．

- n_w 変化前になくかつ n_w 変化後に式 (2) を満たす語が追加される．
- n_w 変化前にはあるが n_w 変化後に式 (2) を満たさなくなる語が削除される．

ここで重要なのは後者の性質である．式 (2) の性質上，後者の変化より W_{ts} から削除された語は n_w をどんなに大きくしても再度 W_{ts} に追加されることは絶対にない． n_w 変化後に削除される語は必ずしも重要度が高い語とも低い語とも言えないが，少なくとも一度は W_{ts} の条件を満たすため重要度が高い語を含む可能性は高い．よって， n_w の変化によって変わる各 W_{ts} の集合には他の集合にはない重要度の高い W_{ts} が含まれている可能性が少なからずあるということになる．したがって， n_w の差異により各分類器に長所・短所が生まれ，アンサンブル学習の効果が生まれやすいということが出来る．逆に， C_2 をアンサンブルしない場合， n_w の差異により性能に大きく差がつくと考えられ， n_w をパラメータとした調整は難しいと考えられる．

また n_w を増やせば，それだけ対象語から離れた位置にある語を特徴とすることになるため，少しずつ W_{ts} の信頼性が落ちていくものと考えられるが，これに伴い任意の s に対し $C_2(n_w)$ の s の推定確率 $p(s|\mathbf{x}, n_w)$ も落ちていくと予想される．すると， C_3 の出力を式 (3) を満たす $s_*(x)$ としたが， n_w の増大に従い $C_2(n_w)$ の判定が採用される確率も減少していくと考えられる．よって， n_w の増大は W_{ts} の信頼性の減少を意味するが，同時に $C_2(n_w)$ の判定の採用確率も減ぜられる．このため本手法は n_w の増大に対し頑健なアンサンブル手法であるとも言えることができる．

本手法は Mihalcea の Smoothed Co-training (Mihalcea 2004) および Wang ら Trajectory Based の手法 (Wang and Matsumoto 2004) と類似性を持つ．まず Wang らは文脈の大きさを変えながら複数の Naive Bayes 分類器を作成しているが，提案手法の処理はこれとよく似ている．Wang らの手法は文脈の大きさというパラメータの影響の差による性能差が小さくなることで性能が上がるが見られるが，本手法でも同様の効果が期待できる．また Mihalcea は Co-training の反復過程にて分類器の多数決を適用した結果，反復回数の差による性能差が小さくなりかつ全体的な性能も向上したと報告したが，本手法における分類器の組合せにおいても同様の効果が期待できる．

なお，分類器の組合せのもう一つの単純な方式として推定確率を重みとした重み付き多数決

方式, つまり

$$s_*(x) = \arg \max_s \left[\sum_{n_w} p(s|\mathbf{x}, n_w) \right] \quad (4)$$

が考えられる. ここで記号の意味は式 (3) と同じである. しかし, 式 (4) の方式は事前実験の結果, 式 (3) の方式ほどは良くないことがわかった. これは, n_w の増大に伴い $C_2(n_w)$ の推定確率が低くなることに変わりはないが, 式 (3) と比べ n_w の大きな $C_2(n_w)$ の判定がより重めに考慮されていることが原因と思われる.

最後に本手法唯一のパラメータである n_w の変化の範囲について述べる. 一つの方法としては範囲を設けない, つまり任意の n_w を許すことが考えられるが, 当然ながら n_w を増やすことで計算時間が増加する. また, n_w の増大に伴う分類器の信頼性の減少に対しある程度は頑健であるとはいえ, 限度の存在があり得る. このため n_w の変化の範囲には何らかの閾値を定めるのが妥当と考えられる. 4 節で述べる実験ではパラメータ $n_{\max}$ を定め, $1 \leq n_w \leq n_{\max}$ の範囲で n_w を 1 刻みで変化させるとし, $n_{\max}$ の変化により語義曖昧性解消の性能がどのように変化していくか見ていく.

3.4 まとめ

提案手法を一つのアルゴリズムとして表現すると次のようになる.

- Step 0** n_w を初期値 (= 1) に設定する.
- Step 1** 訓練データ L 中の語義ラベル s が付与された対象語を T_s とする.
- Step 2** $t_s \in T_s$ 前後 n_w 語以内の式 (2) を満たす内容語の表層形を手掛り語 W_{t_s} として獲得する.
- Step 3** ラベルなしデータ U から対象語 t に対する事例 I_{t0} を抽出する.
- Step 4** $i_{t0} \in I_{t0}$ の前後 $n'_w = 2$ 以内に $w_{ts} \in W_{t_s}$ が出現するとき, i_{t0} を手掛り付き事例 $i_{ts1} \in I_{ts1}$ として抽出する.
- Step 5** L を用いて一次分類器 C_1 を作成し i_{ts1} を分類した結果を $c_1(i_{ts1})$ とする.
- Step 6** $c_1(i_{ts1}) = s$ であるとき, i_{ts1} を $i_{ts2} \in I_{ts2}$ とする.
- Step 7** L に I_{ts2} を加え二次分類器 C_2 を作成する.
- Step 8** n_w を $1 \leq n_w \leq n_{\max}$ の範囲で変化させ, Step 1 から 7 まで繰り返す.
- Step 9** Step 8 で得られた C_2 をアンサンブルして最終分類器 C_3 とする.

まず着目すべきは本手法は Step 8 に示した $n_{\max}$ 以外にパラメータが存在しないことである. そして 4 節で示すようにこのパラメータの設定は比較的容易である.

次に留意すべきは 1 節で述べた P, G, I といったパラメータがないにも関わらず, ブートストラッピングの効果が充分に見込めるという点である. このメカニズムは, 上記 Step 2 に示し

た手掛り語 W_{ts} の条件, Step 4 の I_{ts1} の抽出の条件, Step 6 の I_{ts2} 抽出の条件, さらに Step 8, 9 が巧妙に作用しあっていることに基づいている.

最後に注意すべきは, 本手法は Step 0 から 9 までの 1 度の実行だけで, 十分なブートストラッピングが可能であり, 2 回以上の反復を必要としない点である. しかし, Step 4 の I_{ts1} の抽出は必然的に再現率を犠牲にするため, 本手法 1 回の実行で完全な学習ができる訳ではない. 本手法の反復による更なる精度の向上は今後の課題である.

4 実験

今回実験に使用したデータセットは SemEval-2 日本語タスク (Okumura et al. 2010) において配布された訓練データ (語義ラベル), テストデータを含む形態素解析済みコーパス, および語義の定義に用いられた岩波国語辞典 (西尾, 岩淵, 水谷 1994) のデータのみである. ラベルなしデータとしては上述の形態素解析済みコーパス全てを利用した. 本研究では形態素の基本形の情報も利用するが, SemEval-2 日本語タスクにおいてはデータセットに漢字表記の基本形の情報も付与されていない. しかし, 用言については基本形のカナ表記 (bfm; 語形) の情報が付与されているため, 本研究ではこれを基本形として扱った. また本研究では, 配布されたコーパスの形態素に付与されている品詞が名詞, 動詞, 形容詞, 形状詞, 副詞, 接尾辞, 接頭辞のいずれかであるものを内容語とした.

本研究では, 訓練データに対する前処理として, SemEval-2 日本語タスクにおいて語義の定義に用いられた岩波国語辞典の語釈文に含まれる用例の訓練データへの追加を行った. まず岩波国語辞典の用例中の対象語は “—” に置きかえられているため, “—” を人手で対象語に置換し直し, 用例を自動と人手による修正により抽出した. 用例の形態素解析には SemEval-2 のデータセットの形態素解析に用いられたのと同様に辞書として UniDic を用い, MeCab⁸ を用いて形態素解析した. 形態素解析結果に対する対象語の語義の付与は, 用例の抽出同様, 自動と人手による修正の組合せより行った. ここで追加した用例の総数は 788 であり, 追加後の訓練データ全体の約 24% を占める.

上記の辞書中の用例とオリジナルのデータセットの間には性質に相違がある. オリジナルのデータセットはまとまった文章で与えられ, 対象語周辺から多くの文脈情報が得られる. 対して, 辞書の用例は短い文で与えられ, 得られる文脈情報は少なく, また用例自体は通常の文章に現れにくい表現の場合がある. 3.1 節にて述べた手掛り語獲得の手法は, 対象語から離れた位置からも手掛り語を獲得することを想定した手法であり, 上述のような性質の異なるデータを併用すると不具合が生じる可能性が考えられる. このため実験では, 手掛り語獲得において, 上

⁸ <http://mecab.sourceforge.net/>

記の辞書用例からの追加データを使用する場合と使用しない場合の二通りの実験を行った。

実験は $n_{\max}$ を変化させつつ, 3.3 節で述べた通りに n_w を $1 \leq n_w \leq n_{\max}$ の範囲で 1 刻みで変化させ最終分類器を作成し, 最終分類器の性能を前述のテストデータを用いて評価した. $n_{\max}$ の値は $1 \leq n_{\max} \leq 100$ の範囲で 1 刻みで変化させた。

まず手掛り語獲得において辞書用例データを使用しない場合の実験結果を図 2 に示す. ここで Final classifier が提案手法の分類精度であり, 1st classifier の結果は 3.2 節にて述べた一次分類器によるテストデータの分類精度を示す. なお一次分類器はベースラインの教師あり手法であることに注意されたい. また, 最終分類器と二次分類器の性能比較のため, n_w を $1 \leq n_w \leq 100$ の範囲で変化させ二次分類器単体でテストデータを分類したときの評価結果を図 3 に示す。

図 2 と図 3 の共通点としては, どの $n_{\max}$, n_w においても最終分類器, 二次分類器ともに一次分類器の結果を上回っていることがわかる. これより本手法の第二段階の時点で信頼性の高いブートストラッピングができていたことが確認できる. 一方で図 3 を見て明らかな通り, 二次分類器は最終分類器と比較してパラメータ (n_w) による精度のバラつきがかなり大きいことがわかる. また最終分類器は $n_{\max} = 50$ 付近から, 二次分類器は $n_w = 60$ 付近から $n_{\max}$, n_w の増加に伴い明白に精度が落ちていくことがわかるが, 二次分類器の精度の落ち方は最終分類器と比べ明らかに大きい. これより 3.3 節で述べたように提案手法が n_w の増大に対し実際に頑健であることもわかる. 二次分類器の最高値は最終分類器と比べると有意に高いと言えるが, そのときの $n_w = 23$ のすぐ近くに大きな谷 ($n_w = 27$) があり, これより二次分類器のパラメータ調整の難しさがうかがえる. 一方で, 最終分類器は特に $35 \leq n_{\max} \leq 50$ の範囲で性能が安定しており, パラメータ調整は比較的容易と言える。

次に手掛り語獲得において辞書用例データを使用した場合の最終分類器の評価結果を図 4, 二次分類器の評価結果を図 5 に示す. 始めに図 4 と図 5 を比較すると, やはり最終分類器の方がパラメータの変化に対し性能が頑健であることがわかる. しかし図 2 と図 4 で最終分類器同士を比較すると, $n_{\max} = 10$ 周辺では辞書用例使用の方がわずかに性能が良いが, $n_{\max}$ がこれより大きくなると辞書用例不使用と比べ性能がやや落ちる. $5 \leq n_{\max} \leq 50$ の範囲における性能の安定性を比較すると, これは明確に辞書用例不使用の方が高いと言える. また図 3 と図 5 で二次分類器同士を比較すると, 辞書用例使用の方は $n_w = 20$ 周辺においてやや安定性が高いが, 有意差があるとは言にくく, いずれにせよこれらは最終分類器より性能の安定性が低い. 結論としては, 辞書用例使用の有無ではっきりと有意差があるとは言えないが, 今回の実験では性能の安定性は辞書用例不使用の方が比較的高いとみることにする. 以後は辞書用例不使用の場合の実験結果のみを示す。

続いて $5 \leq n_w \leq 20$, $20 \leq n_w \leq 35$, $35 \leq n_w \leq 50$, $5 \leq n_w \leq 50$ の四つの範囲別 ($n_{\max}$ も同様) の二次分類器・最終分類器の精度の最高値・最低値・平均値, および最頻出語義選択, 一次分類器の精度を表 1 に示す. ここで最高値, 最低値の右括弧内の値はそのときの n_w , $n_{\max}$ を

図 2 $n_{\max}$ の変化による提案手法分類精度の推移 (手掛け語獲得で辞書用例不使用)

図 3 n_w の変化による二次分類器の分類精度の推移 (手掛け語獲得で辞書用例不使用)

図 4 $n_{\max}$ の変化による提案手法分類精度の推移 (手掛け語獲得で辞書用例使用)

図 5 n_w の変化による二次分類器の分類精度の推移 (手掛け語獲得で辞書用例使用)

示す。最高値、最低値の n_w , $n_{\max}$ が複数ある場合はその中で最も小さい値を示した。

表 1 において特筆すべきはやはり最終分類器の最低値と平均値の高さである。以下全て $5 \leq n_{\max} \leq 50$ (n_w も同じ) の範囲について述べる。まず最低値は 0.44 ポイント二次分類器を上回り、一次分類器を常に 1.56 ポイント上回ることになる。これは二次分類器の平均値よりも高い値である。また最終分類器の平均値は一次分類器を 1.69 ポイント上回る。一方、最高値は二次分類器の方が最終分類器を 0.44 ポイント上回り、最大で一次分類器を 2.24 ポイント上回ることになる。しかし、最高値と最低値の差は二次分類器は 1.12 ポイントあるのに対し、最終分類器はわずか 0.24 ポイントである。よって、パラメータの調整を考えると最終分類器の方が手法として扱いやすい。

最後に一次分類器の精度、 $n_{\max}$ が表 1 に示した $5 \leq n_{\max} \leq 50$ の範囲における最高値・最低

表 1 各分類器の分類精度の比較 (手掛り語獲得で辞書用例不使用)

		最高値		最低値		平均値
最頻出語義選択		68.92				
一次分類器		76.60				
二次分類器	$5 \leq n_w \leq 20$	78.68	(20)	77.84	(13)	78.19
	$20 \leq n_w \leq 35$	78.84	(23)	77.72	(27)	78.20
	$35 \leq n_w \leq 50$	78.04	(41)	77.80	(39)	77.93
	$5 \leq n_w \leq 50$	78.84	(23)	77.72	(27)	78.10
最終分類器	$5 \leq n_{\max} \leq 20$	78.32	(6)	78.16	(7)	78.22
	$20 \leq n_{\max} \leq 35$	78.36	(35)	78.16	(20)	78.28
	$35 \leq n_{\max} \leq 50$	78.40	(38)	78.36	(35)	78.38
	$5 \leq n_{\max} \leq 50$	78.40	(38)	78.16	(7)	78.29

値の精度のときの値の最終分類器の精度, および $5 \leq n_{\max} \leq 50$ の範囲での最終分類器の精度の平均値のそれぞれについて, SemEval-2 日本語タスクの語義曖昧性解消対象語別に求めた結果を表 2 に示す. なお表 2 に示した最高値・最低値は $n_{\max}$ を対象語別に精度が最高値・最低値になるように変化させたときの値ではないため, 表 2 の「子供」の結果のように, 最低値の精度が最高値より高い場合も存在する. また対象語右の括弧内の数字は SemEval-2 における判別すべき語義の総数であり, この数字に付記された“+”はテストデータに新語義の語が含まれていたことを示す. ただし本稿では特に新語義判別を考慮した処理はしておらず, 本実験での一次, 二次, 最終の各分類器の比較に新語義の有無は関係しない.

表 2 において強調してある一次分類器の精度は最終分類器の平均値を上回るもの, 強調してある最終分類器の平均値は一次分類器の精度を 5 ポイント以上上回るものを示している. まず最終分類器の平均値を上回る一次分類器の結果を見てみると, その差は最大で「始める」の 3.7 ポイントと比較的小さい. 一方, 一次分類器の結果を上回る最終分類器の結果を見てみると, 最大で「教える」が 14.3 ポイント向上しているとわかる.

対象語の語義数と精度の変化の関係をしてみると, 例外的に「もの」は比較的語義数が多くて精度の向上もやや大きい, 語義数の多い語は一次分類器と比べ最終分類器の精度が落ちる傾向がある. これは, 手掛り語の条件である式 (2) の存在により, 語義数が多いと有益な手掛り語の獲得が難しくなるためと考えられる. 対処方法の一つとしては語義数により手掛り語の条件を変えることが考えられるが, どのように条件を変えるべきかは難しい問題である.

最後に全体を見てみると, 結局一次分類器と最終分類器の間に差がないものが多いことに気付く. よって提案手法は一部の単語に対し高い効果を持つ手法であるということになる. この原因としては, 本手法で手掛り語として用いたのが内容語の表層形の 1 グラムのみと, 手掛り語が素性としては非常に狭い範囲内に位置するものだったこと, そして手掛り付き事例抽出の

表 2 対象語・分類器別の分類精度の比較（手掛り語獲得で辞書用例不使用）

対象語 (語義数)	一次 分類器	最終分類器			対象語 (語義数)	一次 分類器	最終分類器		
		最高値	最低値	平均値			最高値	最低値	平均値
相手 (3)	82.0	82.0	82.0	82.0	強い (2)	92.0	92.0	92.0	92.0
合う (6)	92.0	90.0	92.0	90.3	手 (7+)	76.0	78.0	78.0	78.0
上げる (9+)	50.0	50.0	48.0	49.8	出る (3)	56.0	58.0	58.0	58.4
与える (3)	62.0	62.0	62.0	62.0	電話 (2)	84.0	82.0	82.0	82.0
生きる (2)	90.0	92.0	92.0	92.0	取る (8)	34.0	32.0	30.0	31.2
意味 (3+)	48.0	56.0	56.0	56.2	乗る (4)	60.0	64.0	60.0	63.1
入れる (3)	72.0	74.0	74.0	73.6	場合 (2)	90.0	90.0	90.0	90.0
大きい (3)	94.0	94.0	94.0	94.0	入る (3+)	50.0	58.0	60.0	58.2
教える (3)	60.0	74.0	74.0	74.3	初め (3)	96.0	96.0	96.0	96.0
可能 (1+)	60.0	62.0	62.0	62.0	始める (3+)	90.0	86.0	88.0	86.3
考える (2)	98.0	98.0	98.0	98.0	場所 (3)	94.0	94.0	94.0	94.0
関係 (3)	96.0	96.0	96.0	96.0	早い (2)	80.0	80.0	80.0	80.0
技術 (2)	82.0	82.0	82.0	82.0	一つ (4)	90.0	90.0	90.0	90.0
経済 (2)	98.0	98.0	98.0	98.0	開く (3)	78.0	78.0	80.0	78.5
現場 (2)	82.0	80.0	78.0	79.3	文化 (2)	96.0	98.0	98.0	98.0
子供 (2)	56.0	64.0	68.0	64.6	外 (2)	100.0	100.0	100.0	100.0
時間 (4)	74.0	78.0	80.0	78.3	前 (3+)	72.0	74.0	74.0	74.2
市場 (3)	54.0	66.0	64.0	65.7	見える (4)	76.0	76.0	74.0	75.8
社会 (3)	86.0	86.0	86.0	86.0	認める (4)	78.0	80.0	78.0	78.6
情報 (2)	84.0	84.0	84.0	84.0	見る (6)	86.0	86.0	86.0	86.0
進める (5)	94.0	94.0	94.0	94.0	持つ (4)	78.0	82.0	78.0	80.8
為る (8)	70.0	74.0	74.0	73.0	求める (3+)	72.0	74.0	74.0	74.0
高い (2)	88.0	88.0	88.0	88.0	もの (6)	86.0	92.0	92.0	91.9
出す (4)	50.0	54.0	54.0	54.0	遣る (5)	96.0	96.0	96.0	96.0
立つ (6)	60.0	58.0	58.0	58.6	良い (5)	38.0	48.0	42.0	45.8
平均						76.60	78.40	78.16	78.29

語義数右の“+”はテストデータに新語義の語が含まれていたことを示す。

条件が対象語前後 2 語以内に手掛り語を含むという厳しい制約であったことが考えられる。ここで後者の原因は手掛り付き事例抽出の条件の $n'_w = 2$ を可変にし、さらに多くの二次分類器を作成してアンサンブルすることで取り除ける可能性がある。一方、前者の原因は単に素性の粒度を荒くしただけでは精度が低下する恐れがある。

以上を踏まえると 3.1 節にて述べたヒューリスティックは部分的に有効と言えるが、汎用性に不安が残る。3.1 節の手法第一段階を他の段階と切り離すとすると、第二段階、第三段階に変更を加えない場合、第一段階の要件は

- (1) 手掛り付き事例を（仮の）ラベルを付与して抽出すること
- (2) 抽出事例は第二段階の分類を助ける手掛りを持つこと

(3) 第三段階にて二次分類器を複数作成するためのパラメータを持つこと

(4) 要件 (2) の手掛りの信頼性と要件 (3) のパラメータの間に相関があること

の四つである。この中で特に難しいと思われるのは要件 (2) であるが、これを実現する方法の一つとしては特徴選択がある。第一段階は狭い意味での特徴選択をしているとも見ることができ、効果的な特徴選択法を利用することで上述の問題への対処、つまりより多くの単語の語義曖昧性解消を実現できる可能性がある。特徴選択を応用した語義曖昧性解消の研究の一つとしては Mihalcea によるものが挙げられる (Mihalcea 2002)。Mihalcea は手法は Senseval-2 English all words task において二位の成績から 5.4 ポイント引き離し最高の成績を得た実績がある。よって、Mihalcea の手法を本手法に応用することで更なる性能の向上を見込める可能性がある。

5 おわりに

本稿では、従来のブートストラッピング法による語義曖昧性解消手法の欠点であるパラメータ調整の難しさに対処するため、パラメータ設定をほぼ不要とするブートストラッピング的半教師あり語義曖昧性解消手法を提案した。この手法は二段階の分類をラベルなしデータに適用するものであり、反復回数を一回に留めるにも関わらず十分な効果があるブートストラッピングを実現した。またラベル付きデータに追加するラベルなしデータの条件を変え、複数の分類器を作成しアンサンブル学習することで唯一のパラメータの調整も容易とする手法を確立した。

本手法の改良の方針としては次の二つが考えられる。

- 本手法を通常のブートストラッピング同様に訓練データの反復追加を可能にし、性能が向上するように拡張する。
- 3.1 節にて述べたヒューリスティックによる分類をより多くの語の語義曖昧性解消に有効な汎用的手法に改良する。

これらはいずれも重要度の高い課題であり、並行して取り組むべき課題であると言える。また、提案した手法は語義曖昧性解消に特化しているが、3.2 節と 3.3 節で述べたような半教師ありのアンサンブル学習の枠組みを異なる問題に適用することも興味深い課題である。

最後に、実現可能性は未知数だが本手法の発展の可能性として、語義曖昧性解消のような分類問題において機械学習を用い人間にとって理解しやすい規則を獲得できるようになれば面白いのではないかと筆者は考えている。これは本稿にて提示したヒューリスティックに基づいた手法そのものを訓練事例とみなし、機械が自動でこの訓練事例に類似し、かつ人間に理解しやすい規則を獲得するというものである。これを実現するには「理解しやすさ」という尺度を定義することから始めなければならないと思われるが、もし実現できれば、従来の機械学習において困難だった獲得した規則の解釈が容易になることが予想される。そうすれば機械学習、またはこれを応用した自然言語処理などの研究の進展が加速するのではないかと期待できる。

謝 辞

本研究に用いた評価用データセットをご準備, ご提供くださいました SemEval-2 日本語タスクの運営の皆様, ならび本稿のために有益なコメントをお寄せくださいました査読者の皆様に深く感謝いたします.

参考文献

- Abney, S. (2004). “Understanding the Yarowsky algorithm.” *Computational Linguistics*, **30** (3), pp. 365–395.
- Agirre, E., Martínez, D., de Lacalle, O. L., and Soroa, A. (2006). “Two Graph-Based Algorithms for State-of-the-art WSD.” In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 585–593.
- Banerjee, S. and Pedersen, T. (2003). “Extended Gloss Overlaps as a Measure of Semantic Relatedness.” In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 805–810.
- Blum, A. and Mitchell, T. (1998). “Combining Labeled and Unlabeled Data with Co-Training.” In *COLT: Proceedings of the Workshop on Computational Learning Theory*, pp. 92–100.
- Brody, S., Navigli, R., and Lapata, M. (2006). “Ensemble Methods for Unsupervised WSD.” In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL)*, pp. 97–104.
- Curran, J. R., Murphy, T., and Scholz, B. (2007). “Minimising Semantic Drift with Mutual Exclusion Bootstrapping.” In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, pp. 172–180.
- Edmonds, P. and Cotton, S. (2001). “SENSEVAL-2: Overview.” In *Proceedings of the 2nd International Workshop on Evaluating Word Sense Disambiguation Systems (Senseval-2)*, pp. 1–6.
- Escudero, G., Màrquez, L., and Rigau, G. (2000). “Boosting Applied to Word Sense Disambiguation.” In *Proceedings of the 11th European Conference on Machine Learning (ECML)*, pp. 129–141.
- Florian, R., Cucerzan, S., Schafer, C., and Yarowsky, D. (2002). “Combining classifiers for word sense disambiguation.” *Natural Language Engineering*, **8** (4), pp. 327–431.
- Fujita, S., Duh, K., Taira, H., and Shindo, H. (2010). “MSS: Investigating the Effectiveness of

- Domain Combinations and Topic Features for Word Sense Disambiguation.” In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, pp. 383–386.
- Galley, M. and McKeown, K. (2003). “Improving Word Sense Disambiguation in Lexical Chaining.” In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1486–1488.
- Joachims, T. (2003). “Transductive Learning via Spectral Graph Partitioning.” In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pp. 290–297.
- Klein, D., Toutanova, K., Ilhan, H. T., Kamvar, S. D., and Manning, C. D. (2002). “Combining Heterogeneous Classifiers for Word-Sense Disambiguation.” In *Proceedings of the ACL workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pp. 74–80.
- 小町守, 工藤拓, 新保仁, 松本裕治 (2010). Espresso 型ブートストラッピング法における意味ドリフトのグラフ理論に基づく分析: 語義曖昧性解消における評価. 人工知能学会論文誌, **25** (2), pp. 233–242.
- Le, A.-C., Shimazu, A., Huynh, V.-N., and Nguyen, L.-M. (2008). “Semi-supervised learning integrated with classifier combination for word sense disambiguation.” *Computer Speech and Language*, **22** (4), pp. 330–345.
- Liu, D. C. and Nocedal, J. (1989). “On the limited memory BFGS method for large scale optimization.” *Mathematical programming*, **45**, pp. 503–528.
- Màrquez, L., Escudero, G., Martínez, D., and Rigau, G. (2004). “Supervised Corpus-Based Methods for WSD.” In Agirre, E. and Edmonds, P. (Eds.), *Word Sense Disambiguation: Algorithms and Applications*, pp. 167–216. Springer.
- McCarthy, D., Koeling, R., Weeds, J., and Carroll, J. (2004). “Finding Predominant Word Senses in Untagged Text.” In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 280–287.
- Mihalcea, R. (2002). “Word sense disambiguation with pattern learning and automatic feature selection.” *Natural Language Engineering*, **8** (4), pp. 343–358.
- Mihalcea, R. (2004). “Co-training and Self-training for Word Sense Disambiguation.” In *Proceedings of the 8th Conference on Computational Natural Language Learning (CoNLL)*, pp. 33–40.
- Mihalcea, R. and Edmonds, P. (Eds.) (2004). *Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3)*.
- Navigli, R. (2009). “Word sense disambiguation: A survey.” *ACM Computing Surveys*, **41** (2), pp. 1–69.
- Navigli, R. and Velardi, P. (2005). “Structural semantic interconnections: A knowledge-based ap-

- proach to word sense disambiguation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27** (7), pp. 1075–1088.
- Nigam, K. and Ghani, R. (2000). “Analyzing the Effectiveness and Applicability of Cotraining.” In *Proceedings of the 9th International Conference on Information and Knowledge Management*, pp. 86–93.
- Nigam, K., Laerty, J., and McCallum, A. (1999). “Using Maximum Entropy for Text Classification.” In *Proceedings of IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp. 61–67.
- 西尾実, 岩淵悦太郎, 水谷静夫 (1994). 岩波国語辞典 第五版. 岩波書店.
- Niu, Z.-Y., Ji, D.-H., and Tan, C. L. (2005). “Word Sense Disambiguation Using Label Propagation Based Semi-Supervised Learning.” In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 395–402.
- Okumura, M., Shirai, K., Komiya, K., and Yokono, H. (2010). “SemEval-2010 Task: Japanese WSD.” In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, pp. 69–74.
- Pantel, P. and Pennacchiotti, M. (2006). “Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations.” In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pp. 113–120.
- Pedersen, T. (2000). “A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation.” In *Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 63–69.
- Pham, T. P., Ng, H. T., and Lee, W. S. (2005). “Word Sense Disambiguation with Semi-Supervised Learning.” In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI)*, pp. 1093–1098.
- 新納浩幸 (2001). 素性間の共起性を検査する Co-training による語義判別規則の学習. 情報処理学会 自然言語処理研究会 NL-145-5, pp. 29–36.
- 鶴岡慶雅, 近山隆 (2002). ベイズ統計の手法を利用した決定リストのルール信頼度推定法. 自然言語処理, **9** (3), pp. 3–19.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley and Sons.
- Véronis, J. (2004). “HyperLex: Lexical cartography for information retrieval.” *Computer Speech and Language*, **18** (3), pp. 223–252.
- Wang, X. and Matsumoto, Y. (2004). “Trajectory Based Word Sense Disambiguation.” In *Proceedings of the 20th international conference on Computational Linguistics*, pp. 903–909.

八木豊, 野呂智哉, 白井清昭, 徳永健伸, 田中穂積 (2001). 決定リストを用いた語義曖昧性解消. 電子情報通信学会 言語とコミュニケーション研究会 NLC2001-42, pp. 47–52.

Yarowsky, D. (1995). “Unsupervised Word Sense Disambiguation Rivaling Supervised Methods.” In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pp. 189–196.

Zhou, Z.-H. and Li, M. (2005). “Tri-training: Exploiting Unlabeled Data using Three Classifiers.” *IEEE Transactions on Knowledge and Data Engineering*, **17** (11), pp. 1529–1541.

Zhu, X. and Ghahramani, Z. (2002). “Learning from Labeled and Unlabeled Data with Label Propagation.” Tech. rep., CMU CALD tech report CMU.CALD.02.107.

略歴

井上 裁都：2009 年慶應義塾大学理工学部情報工学科卒業. 2011 年同大学院修士課程修了. 修士（工学）.

斎藤 博昭：慶應義塾大学理工学部情報工学科准教授. 工学博士.

(2010 年 9 月 30 日 受付)

(2010 年 11 月 9 日 採録)