合議による語義曖昧性解消の領域適応のための確信度の調整

古宮　嘉那子（東京農工大学　工学研究院）†
奥村　学（東京工業大学　精密工学研究所）
小谷　善行（東京農工大学　工学研究院）

Adjustment of Confidence Score for Domain Adaptation in Word Sense Disambiguation Based on the Comparison of Multiple Classifiers

Kanako Komiya (Institution of Engineering, Tokyo University of Agriculture and Technology)
Manabu Okumura (Precision and Intelligence Laboratory, Tokyo Institution of Technology)
Yoshiyuki Kotani (Institution of Engineering, Tokyo University of Agriculture and Technology)

１．はじめに

テストのターゲットとなるドメインとは異なるドメインのデータ（ソースデータ）を利用して学習を行い，ターゲットドメインのデータ（ターゲットデータ）に適応することを領域適応といい，近年さまざまな手法が研究されている．

（Komiya and Okumura (2012)）では，語義曖昧性解消（Word Sense Disambiguation，WSD）の領域適応の適切な手法は用例によって異なると考え，supervised な領域適応において，分類器の出力する確信度の高い方の答えを採用することにより，分類の精度を向上させる手法を示した．また，（古宮，奥村，小谷 (2013)）では，同様の手法が unsupervised な領域適応においても効果を発揮することを示した．しかし，この手法は，分類器ごとの訓練事例数に大きなばらつきがある際には効果が上がらないという問題があった．そのため，本稿では，合議による語義曖昧性解消の領域適応のために分類器間の確信度を調整することを提案する．unsupervised な領域適応において，確信度に対してコーパス中に出現した語義数により調整を行った結果，正解率に有意な上昇が見られた．

２．関連研究

領域適応は，学習に使用する情報により，supervised，semi-supervised，unsupervised の三種に分けられる．本研究で扱うのは，unsupervised の領域適応，つまりラベルつきのソースデータとラベルなしのターゲットデータを利用するものである．

（古宮，奥村 (2012)）は WSD について supervised な領域適応を行った場合，最も効果的な領域適応手法はソースデータとターゲットデータの性質により異なることを示し，最も効果的な領域適応手法を，WSD の対象単語タイプ，ソースデータ，ターゲットデータの三つ組ごとに自動的に選択する手法について述べた．

また，（Komiya and Okumura (2012)）は，WSD の supervised な領域適応において，本稿でも使用する確信度という尺度を用い，用例ごとに訓練事例を自動的に選択した．（古宮，奥村，小谷 (2013)）は（Komiya and Okumura (2012)）の手法が unsupervised の領域適応に対しても有効であることを示している．本稿はこの（古宮，奥村，小谷 (2013)）についての改良を行う．

最後に，（古宮，小谷，奥村 (2013)）は，unsupervised な領域適応において，あるターゲットデータに対して複数のジャンルのソースデータが混在した場合，確信度と LOO-bound という指標を利用して，領域適応のための訓練事例の部分集合を WSD の対象単語タイプごとに自動的に選択する手法について述べた．

†kkomiya@cc.tuat.ac.jp

３．合議のための確信度の調整

あるドメイン（ジャンル）のターゲットデータを対象に WSD を行うことを考える．このターゲットデータのラベル（語義）は未知であるとする．一方，複数のジャンルのコーパスの集合となっているソースデータが入手可能であるとすると，本稿ではこれらのソースデータの全体集合から，ターゲットデータに適した訓練事例の部分集合を自動的に選択する．この際，以下の手順で訓練事例の部分集合の選択を行う．なお，我々は最も効果的な訓練事例集合は用例ごとに異なると仮定しているため，訓練事例集合の選択はターゲットデータの用例ごとに行う．

(1) 訓練事例集合を変えて複数の分類器を学習する．
(2) 用例ごとに，複数の訓練事例集合による分類器の確信度を比較する．
(3) 分類器の確信度の最も高い訓練事例集合による結果を採用する．

ここでの分類器のスコアには確信度（Komiya and Okumura (2012)）を利用し，その値を比較することで，複数の分類器の合議を行う．ここで，確信度とは分類の確からしさを表す 0〜1 の値であり，active-learning においてラベル付けする用例を選択するのによく利用される．

しかし，（古宮，小谷，奥村　(2013)）で示されたように，この手法は，分類器ごとの訓練事例に大きなばらつきがある際には効果が上がらないという問題があった．これは，個々の分類器の訓練事例のみを基準として出力される確信度という値を，分類器間の出力結果を比較するために用いているためである．そのため本稿では，分類器間の比較に耐えるように確信度を調整する手法として，以下の三つを提案し，比較する．

・分類器が出力した語義の，訓練事例における事前確率でスコアを割る（Prior）．
・分類器の訓練事例における最頻出語義の事前確率でスコアを割る（MFS）．
・分類器の訓練事例に出現する語義数をスコアにかける(SNum).

Prior は，事前確率が高い語義を回答する分類器と，低い語義を答える分類器では，同じ確信度でも，事前確率の低い語義を答える分類器をより評価するべきだとする手法である．また MFS は同様に，最頻出語義の事前確率が低いデータで学習した分類器の確信度の方を，最頻出語義の事前確率が高いデータで学習した分類器の確信度よりも評価すべきだとする手法である．最後に SNum は，訓練事例に現れる語義が多いデータで学習した分類器の確信度の方を，訓練事例に現れる語義が少ないデータで学習した分類器よりも評価すべきだとする手法である．

４．実験

分類器としてはマルチクラス対応の SVM（libsvm）(Chang and Lin (2001))を使用した．また，libsvm の確率として出力される分類の確からしさを確信度として用いた．カーネルは予備実験の結果，線形カーネルが最も高い正解率を示したため，これを採用した．また，WSD の素性には，以下の 17 種類の素性を用いた．

・WSD の対象単語の前後二語までの形態素の表記（4 種類）
・WSD の対象単語の前後二語までの品詞（4 種類）
・WSD の対象単語の前後二語までの品詞の細分類（4 種類）
・WSD の対象単語の前後二語までの分類コード（4 種類）
・ 係り受け（1 種類）
　- 対象単語が名詞の場合はその名詞が係る動詞
　- 対象単語が動詞の場合はその動詞のヲ格の格要素

分類語彙表の分類コードには（国立国語研究所 (1964)）を使用した.
また，実験は五分割交差検定を用いた.
また，本研究では二つのコーパスが与えられたと仮定し，それぞれのコーパスを使った場合と，さらに入手可能な二つのコーパスをすべて使った場合の三通りのうちから WSD のための訓練事例集合を選択した．また，合議の際には，最も高い確信度の分類器の結果（語義）を採用した.

５．実験データ

実験には，現代日本語書き言葉均衡コーパス（BCCWJ コーパス）(Maekawa (2008)) の Yahoo!知恵袋，白書，Yahoo！ブログ，書籍，雑誌，新聞のコアデータ[1]と，コアデータではない白書と Yahoo! 知恵袋のデータ，また RWC コーパスの毎日新聞コーパス（Hashida et al. (1998)）の合計 9 つのデータを利用した．これらのデータには岩波国語辞典（西尾ら (1994)) の語義が付与されている. 九つのコーパスのうち, ひとつをターゲットデータにし，残りのうちの二つを利用可能なソースデータとして利用して領域適応の実験を行った．この際，使用した三つのコーパス中全てに 50 トークン以上存在する単語を実験対象とした. 白書と Yahoo! 知恵袋の重複利用を避け, また実験対象となる単語が存在しない組み合わせを除くと，コーパスの選び方は全部で 171 通りとなった．また，対象単語の種類数はコーパスの選び方により変わるが，全ての論理和をとると，49 種類となった.
それぞれのデータにおける 50 用例以上の単語数, 単語ごとの用例数の平均と標準偏差を表 1 に示す.

表 1　それぞれのデータにおける 50 用例以上の単語数，単語ごとの用例数の平均と標準偏差

コーパスの種類	単語	平均	標準偏差
コア　Yahoo！知恵袋	22	157.77	153.77
コア　白書	5	79.20	21.01
コア　Yahoo！ブログ	9	245.22	431.84
コア　書籍	35	158.91	204.97
コア　雑誌	26	284.92	872.53
コア　新聞	25	92.28	78.08
非コア　白書	38	2,088.84	2,234.52
非コア　Yahoo！知恵袋	42	3,979.17	5,786.02
RWC　新聞	66	473.79	1,844.40

また，実験には岩波国語辞典の中分類の語義を採用した．語義数ごとの単語の内訳は，2 語義 :「一般」,「書く」,「考える」,「技術」,「経済」,「現在」,「子供」,「自分」,「情報」,「高い」,「作る」,「強い」,「電話」,「場合」,「他」，3 語義 :「言う」,「今」,「入れる」,「買う」,「関係」,「聴く」,「社会」,「進む」,「地方」,「出来る」,「出る」,「入る」,「開く」,「前」,

[1]コアデータは形態素解析の結果を人手で訂正しているため正確であるが，量が少ない.

「求める」，4 語義：「訴える」，「時間」，「時代」，「出す」，「乗る」，「計る」，「ひとつ」，「見える」，「持つ」，5 語義：「進める」，「やる」，「良い」，6 語義：「会う」，「見る」，「もの」，7 語義：「手」，8 語義：「する」，「取る」，9 語義：「上げる」である．

６．結果

表 2 に使用したデータの区分と適応手法別の実験結果を示す．このうちの「9 つのコーパス」が 9 つのコーパス全ての結果の平均である．また，「コアデータ」には 6 つの BCCWJ のコアデータの結果の平均を，「コアデータ以外」には BCCWJ のふたつのコアデータ以外のデータと RWC コーパスの結果の平均を示した．

表において，「Self」は，タグつきターゲットデータが手に入ったと仮定して，supervised の学習を 5 分割交差検定を用いて行った結果である．「平均的なコーパス」は，ふたつのジャンルのソースデータそれぞれをジャンルごとに分けて訓練事例とした場合の結果の平均である．入手可能なコーパスをそれぞれソースデータとして使用した場合の平均的な結果を示している．例えば，Yahoo！知恵袋のデータがターゲットデータの時のソースデータが白書と新聞だった場合，このときの「平均的なコーパス」は，白書の全データで訓練した Yahoo！知恵袋のデータの正解率と，新聞の全データで訓練した Yahoo！知恵袋のデータの正解率の平均となる．最後に，「全てのコーパス」とは，ふたつのジャンルのソースデータ全て（つまり全ソースデータ）を訓練事例とした際の結果である．例えば，上記の例において，Yahoo！知恵袋のデータがターゲットデータの時の「全てのコーパス」は，白書と新聞のコーパス全てを訓練事例として利用した際の結果である．

このとき，「Self」は upper bound であり，「平均的なコーパス」，「全てのコーパス」はベースラインである．表において，これらのうち Self 以外でコーパスごとに一番高い正解率を太字で示した．

表 2　コーパスと適応手法別の実験結果

データの種類	9 つのコーパス		コアデータ以外		コアデータ	
手法	マイクロ	マクロ	マイクロ	マクロ	マイクロ	マクロ
Self	92.35%	86.44%	91.89%	85.15%	90.07%	85.21%
平均的なコーパス	71.33%	75.47%	80.53%	79.07%	68.77%	77.53%
全てのコーパス	74.60%	77.97%	84.57%	83.31%	79.27%	80.92%
確信度	74.59%	77.92%	83.00%	81.85%	77.34%	79.85%
確信度 Prior	70.86%	75.98%	83.71%	82.06%	66.10%	77.16%
確信度 MFS	72.22%	77.33%	83.62%	81.76%	69.84%	79.26%
確信度 Snum	74.85%	78.13%	84.68%	82.51%	77.54%	80.23%

７．考察

まず，表 2 を見てみると，9 つのコーパス全ての平均では，マイクロ平均でもマクロ平均でも確信度に語義数をかけて調整する手法，「SNum」が最も良いことが分かる．コアデータ以外のデータの平均では，マイクロ平均では「SNum」が最も良いが，マクロ平均ではベースラインである「全てのコーパス」，コアデータの平均ではマクロ平均でもマイクロ平均でも「全てのコーパス」が最も良い．また，「Prior」，「MFS」の調整は効果がなかった．

一方，表 1 を見てみると，コアデータは用例数の平均がコアデータ以外のデータに比べて低いことが分かる．確信度には分類器間で訓練事例数に大きなばらつきが出た場合に効果があがりにくいという欠点があったが，これは訓練事例数が少ない際に，確信度の正確性が低くなるためである．たとえば，訓練集合の用例数が 1 件になると，他に選択肢がないため，その分類器の確信度は最高値の 1 となる．しかし 1 件の訓練事例による分類器が最も良い分類器であるとは考え難く，訓練事例数の少ない場合の確信度はあまり信用できないことが分かる．このように，確信度は訓練事例数が少ないと，多い場合よりも高くなる場合がある．これは直感に反した現象であり，少ない訓練事例数の分類器を比較対象に含める場合には，確信度は不利であると言える．このことから，コアデータは訓練事例数が少ないため，確信度の正確性が低くなり，「全てのコーパス」に及ばなかったと考えられる．

また同じく表 1 から，コアデータ以外のデータは標準偏差がコアデータに比べて非常に高いことが分かる．このことから，コアデータ以外のデータでは，コアデータよりも分類器間で訓練事例数にばらつきが出る傾向が見て取れる．また，コアデータとそれ以外のデータではもっとその傾向が強まると思われる．そのため，9 つのコーパスの平均とコアデータ以外のデータの平均では，分類器間の整合性を調整する手法が効いたと考えられる．なお，9 つのコーパスマイクロ平均での「SNum」と「全てのコーパス」の差は有意水準は 0.05 のとき，カイ二乗検定で有意であった．これらの結果から，確信度は訓練データの少ない分類器においてはその正確性が薄れるため，訓練事例数が少ない二つのコーパスに対しては全データを利用した方が WSD の正解率が高くなること，また，訓練事例数の少ないコーパスと訓練事例数が多いコーパスを比較する場合には，確信度の整合性を語義数で調整することで WSD の正解率が上昇することが分かった．

また表 2 において，先行研究（古宮，奥村，小谷　(2013)）や（Komiya and Okumura (2012)）の結果に反して，ベースラインである「全てのコーパス」が強く，調整を行わない，確信度単独による手法よりも正解率が高くなっているが，これはコアデータにおいては，使用できる訓練事例数が少なくなったためであり，またそれ以外のデータにおいては分類器間の訓練事例数の差が増したためであると考えられる．

８．まとめ

本稿では，unsupervised な領域適応において，あるターゲットデータに対して複数のジャンルのソースデータが混在した場合，用例ごとに最もよい訓練事例集合で学習された分類器を，分類器の確信度を指標として選択する手法を改良した．分類器間の確信度を，コーパス中に出現した語義数により調整を行った結果，正解率に有意な上昇が見られた．

謝　辞

本研究は，文部科学省科学研究費補助金[若手 B（No : 24700138）]の助成により行われた．ここに，謹んで御礼申し上げる．

文献

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Koichi Hashida, Hitoshi Isahara, Takenobu Tokunaga, Minako Hashimoto, Shiho Ogino, and Wakako Kashino (1998). “The Rwc Text Databases”. In *Proceedings of The First International Conference on Language Resource and Evaluation*, pp. 457–461.

Kanako Komiya and Manabu Okumura (2012). Automatic Domain Adaptation for Word Sense Disambiguation Based on Comparison of Multiple Classifiers, *Proceedings of 26th Pacific Asia Conference on Language Information and Computation*, pp 77-85.

Kikuo Maekawa (2008). “Balanced Corpus of Contemporary Written Japanese”. In *Proceedings of the 6th Workshop on Asian Language Resources (ALR)*, pp. 101–102.

国立国語研究所（1964). “分類語彙表”. 秀英出版.

古宮嘉那子，奥村学（2012). “語義曖昧性解消のための領域適応手法の決定木学習による自動選択”，自然言語処理，Vol.19, No.3, pp.143-166.

古宮嘉那子，小谷善行，奥村学（2013). “語義曖昧性解消の領域適応のための訓練事例集合の選択”，第19回言語処理学会年次大会予稿集，pp.940-943.

古宮嘉那子，奥村学，小谷善行(2013). 分類器の確信度を用いた合議制による語義曖昧性解消の semi-supervised な領域適応. 第三回コーパス日本語学ワークショップ予稿集, pp. 1-6.

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.