共変量シフトの問題としての語義曖昧性解消の領域適応

新納　浩幸[†]・佐々木　稔[†]

本稿では語義曖昧性解消 (Word Sense Disambiguation, WSD) の領域適応が共変量シフトの問題と見なせることを示し，共変量シフトの解法である確率密度比を重みにしたパラメータ学習により，WSD の領域適応の解決を図る．共変量シフトの解法では確率密度比の算出が鍵となるが，ここでは Naive Bayes で利用されるモデルを利用した簡易な算出法を試みた．そして素性空間拡張法により拡張されたデータに対して，共変量シフトの解法を行う．この手法を本稿の提案手法とする．BCCWJ コーパスの 3 つ領域 OC （Yahoo! 知恵袋），PB（書籍）及び PN（新聞）を選び，SemEval-2 の日本語 WSD タスクのデータを利用して，多義語 16 種類を対象に，WSD の領域適応の実験を行った．実験の結果，提案手法は Daumé の手法と同等以上の正解率を出した．本稿で用いた簡易な確率密度比の算出法であっても共変量シフトの解法を利用する効果が高いことが示された．より正確な確率密度比の推定法を利用したり，最大エントロピー法の代わりに SVM を利用するなどの工夫で更なる改善が可能である．また教師なし領域適応へも応用可能である．WSD の領域適応に共変量シフトの解法を利用することは有望であると考えられる．

キーワード：語義曖昧性解消，領域適応，共変量シフト，Daumé の手法，BCCWJ コーパス

Domain Adaptations for Word Sense Disambiguation under the Problem of Covariate Shift

Hiroyuki Shinnou[†] and Minoru Sasaki[†]

In this report, we show that the problem of domain adaptation for word sense disambiguation (WSD) can be treated as a covariate shift problem, and we try to solve it by maximizing the log-likelihood by weighting the probability density ratio, which is the standard solution of covariate shift. The key to solving this problem lies in the estimation of the probability density ratio. We estimate the probability density ratio using simple method employing the Naive Bayes model. In our proposed method, we apply the covariate shift method to the training data expanded by the Daumé's feature augmentation method. In the experiment, we solve six types of domain adaptations for WSD using three domains, viz., OC (Yahoo! Chiebukuro), PB (Book), and PN (Newspaper) in the BCCWJ corpus. The results show that our proposed method outperforms the Daumé's method. This report shows that even our simple method of estimating the probability density ratio is effective for use in the covariate shift method. In future, we intend to investigate and find a method of estimating the probability density ratio more accurately. Further, we intend to use the SVM instead

[†] 茨城大学工学部情報工学科, Department of Computer and Information Sciences, Ibaraki University

自然言語処理 Vol. 21 No. 1 March 2014

of the maximum entropy method. Moreover, the method of covariate shift is also effective for unsupervised domain adaptations and is a promising approach for WSD domain adaptations.

Key Words: *word sense disambiguation, domain adaptation, covariate shift, Daumé's method, BCCWJ corpus*

1 はじめに

本稿では語義曖昧性解消 (Word Sense Disambiguation, WSD) をタスクとした領域適応の問題が共変量シフトの問題と見なせることを示す．そして共変量シフトの解法である確率密度比を重みにしたパラメータ学習により，WSD の領域適応の解決を図る．共変量シフトの解法では確率密度比の算出が鍵となるが，ここでは Naive Bayes で利用されるモデルを利用した簡易な算出法を試みた．そして素性空間拡張法により拡張されたデータに対して，共変量シフトの解法を行う．この手法を本稿の提案手法とする．

自然言語処理の多くのタスクにおいて帰納学習手法が利用される．そこではコーパス S からタスクに応じた訓練データを作成し，その訓練データから分類器を学習する．そしてこの分類器を利用することで当初のタスクを解決する．このとき実際のタスクとなるデータはコーパス S とは領域が異なるコーパス T のものであることがしばしば起こる．この場合，コーパス S（ソース領域）から学習された分類器では，コーパス T （ターゲット領域）のデータを精度良く解析することができない問題が生じる．これが領域適応の問題であり[1]，近年活発に研究が行われている (Sogaard 2013).

WSD は文 $\boldsymbol{x}$ 内の多義語 w の語義 $c \in C$ を識別する問題である．$P(c|\boldsymbol{x})$ を文 $\boldsymbol{x}$ 内の単語 w の語義が c である確率とすると，確率統計的には $\arg\max_{c \in C} P(c|\boldsymbol{x})$ を解く問題といえる．例えば単語 $w =$「ボタン」には少なくとも c_1 : 服のボタン，c_2 : スイッチのボタン，c_3 : 花のボタン（牡丹），の 3 つの語義がある．そして文 $\boldsymbol{x} =$「シャツのボタンが取れた」が与えられたときに，文中の「ボタン」が $C = \{c_1, c_2, c_3\}$ 内のどれかを識別する．直接的には教師付き学習手法を用いて $P(c|\boldsymbol{x})$ を推定して解くことになる．WSD の領域適応の問題は，前述したように，教師付き学習手法を利用する際に学習もとのソース領域のコーパス S と，分類器の適用先であるターゲット領域のコーパス T が異なる問題である．領域適応ではソース領域 S から S 上の条件付き分布 $P_S(c|\boldsymbol{x})$ は学習できるという設定なので，$P_S(c|\boldsymbol{x})$ やその他の情報を利用して，ターゲット領域 T 上の条件付き分布 $P_T(c|\boldsymbol{x})$ を推定できれば良い．ここで「シャツのボタンが取れた」という文中の「ボタン」の語義は，この文がどのような領域のコーパスに現れても変化するとは考えづらい．つまり $P_T(c|\boldsymbol{x})$ は領域に依存していないため，$P_S(c|\boldsymbol{x}) = P_T(c|\boldsymbol{x})$ が成立していると

[1] 領域適応は機械学習の分野では転移学習 (神嶌 2010) の一種と見なされている．

考えられる．今 $P_S(c|\boldsymbol{x})$ は推定できるので，$P_S(c|\boldsymbol{x}) = P_T(c|\boldsymbol{x})$ が成立していれば，$P_T(c|\boldsymbol{x})$ を推定する必要はないように見える．ただしソース領域だけを使って推定した $P_S(c|\boldsymbol{x})$ では，実際の識別精度は低い場合が多い．それは $P_S(\boldsymbol{x}) \neq P_T(\boldsymbol{x})$ から生じている．$P_S(c|\boldsymbol{x}) = P_T(c|\boldsymbol{x})$ だが $P_S(\boldsymbol{x}) \neq P_T(\boldsymbol{x})$ という仮定の下で，$P_T(c|\boldsymbol{x})$ を推定する問題は共変量シフトの問題 (Shimodaira 2000; 杉山 2006; Sugiyama and Kawanabe 2011) である．本稿では WSD の領域適応の問題を共変量シフトの問題として捉え，共変量シフトの解法を利用して WSD の領域適応を解決することを試みる．

訓練データを $D = \{(\boldsymbol{x_i}, c_i)\}_{i=1}^N$ とする．共変量シフトの標準的な解法では $P_T(c|\boldsymbol{x})$ に確率モデル $P(c|\boldsymbol{x};\boldsymbol{\theta})$ を設定し，次に確率密度比 $r(\boldsymbol{x_i}) = P_T(\boldsymbol{x_i})/P_S(\boldsymbol{x_i})$ を重みにした以下の対数尤度を最大にする $\boldsymbol{\theta}$ を求めることで，$P_T(c|\boldsymbol{x})$ を構築する．

$$\sum_{i=1}^{N} r(\boldsymbol{x_i}) \log P(c_i|\boldsymbol{x_i};\boldsymbol{\theta})$$

また領域適応に対しては Daumé の手法 (Daumé 2007) が非常に簡易でありながら，効果が高い手法として知られている．Daumé の手法は，データの表現を領域適応に効果が出るように拡張し，拡張されたデータを用いて SVM 等の学習手法を利用する手法である．ここでは拡張する手法を「素性空間拡張法 (Feature Augmentation)」と呼び，拡張されたデータを用いて SVM などで識別までを行う手法を「Daumé の手法」と呼ぶことにする．拡張されたデータに対しては任意の学習手法が利用できる．つまり素性空間拡張法により拡張されたデータに対して，共変量シフトによる解法を利用することも可能である．本稿ではこの手法を提案手法とする．

実験では現代日本語書き言葉均衡コーパス（BCCWJ コーパス (Maekawa 2007)）における 3 つの領域 OC （Yahoo! 知恵袋），PB （書籍）及び PN （新聞）を利用する．SemEval-2 の日本語 WSD タスク (Okumura, Shirai, Komiya, and Yokono 2010) ではこれらのコーパスの一部に語義タグを付けたデータを公開しており，そのデータを利用する．すべての領域である程度の頻度が存在する多義語 16 単語を対象にして，WSD の領域適応の実験を行う．領域適応としては OC → PB，PB → PN，PN → OC，OC → PN，PN → PB，PB → OC の計 6 通りが存在する．結果 $16 \times 6 = 96$ 通りの WSD の領域適応の問題に対して実験を行った．その結果，提案手法は Daumé の手法と同等以上の正解率を出した．

本稿で用いた簡易な確率密度比の算出法であっても共変量シフトの解法を利用する効果が高いことが示された．より正確な確率密度比の推定法を利用したり，SVM を利用するなどの工夫で更なる改善が可能である．また教師なし領域適応へも応用可能である．WSD の領域適応に共変量シフトの解法を利用することは有望であると考えられる．

2 関連研究

自然言語処理における領域適応は，帰納学習手法を利用する全てのタスクで生じる問題であるために，その研究は多岐にわたる．利用手法をおおまかに分類すると，ターゲット領域のラベル付きデータを利用するかしないかで分類できる．利用する場合を教師付き領域適応手法，利用しない場合を教師なし領域適応手法と呼ぶ．本稿における手法は教師付き領域適応手法の範疇に入るので，ここでは提案手法に関連する教師付き領域適応手法の従来研究を述べる．

教師付き領域適応手法においては，一般に，ターゲット領域の知識は使えるだけ使えばよいはずなので，ポイントはソース領域の知識の利用方法にある．ソース領域とターゲット領域間の距離が離れすぎている場合，ソース領域の知識を使いすぎると分類器の精度が悪化する現象がおこる．これは負の転移 (Rosenstein, Marx, Kaelbling, and Dietterich 2005) と呼ばれている．負の転移を避けるには，本質的に，ソース領域とターゲット領域間の距離を測り，その距離を利用してソース領域の知識の利用を制御する形となる．

Asch は品詞タグ付けをタスクとして領域間の類似性を測り，その類似度から領域適応を行った際に精度がどの程度悪くなるかを予測できることを示した (Van Asch and Daelemans 2010)．張本は構文解析をタスクとしてターゲット領域を変化させたときの精度低下の要因を調査し，そこから新たな領域間の類似性の尺度を提案している (張本，宮尾，辻井 2010)．Plank は構文解析をタスクとして領域間の類似性を測ることで，ターゲット領域を解析するのに最も適したソース領域を選んでいる (Plank and van Noord 2011)．Ponomareva (Ponomareva and Thelwall 2012) や Remus (Remus 2012) は感情極性分類をタスクとして領域間の類似度を学習中のパラメータに利用した．これらの研究はタスク毎に類似性を測るが，WSD がタスクの場合，領域間の類似性は WSD の対象単語に依存していると考えられる．古宮は対象単語毎に領域間の距離を含めた性質[2]によって適用する学習手法を変化させている (Komiya and Okumura 2011, 2012; 古宮，奥村 2012)．

上記した古宮の一連の研究は広い意味でアンサンブル学習の一種である．そこでアンサンブルされる各要素となる学習手法をみるとソース領域のデータとターゲット領域のデータへの各重みが異なるだけである．つまり領域適応においてはソース領域のデータとターゲット領域のデータへの各重みを調整して，学習手法を適用するというアプローチが有力である．Jiang は $P_S(c|\boldsymbol{x})$ と $P_T(c|\boldsymbol{x})$ との差が極端に大きいデータを “misleading” データとして訓練データから取り除いて学習することを試みた．これは “misleading” データの重みを 0 にした学習と見なせるため，この手法も重み付けの手法と見なせる．本稿で利用する共変量シフト下での学習もこの範疇の手法といえる．

[2] これら性質を全て含めて，領域間の類似性と呼べる．

素性空間拡張法 (Daumé 2007) も重み付け手法である．ただしデータではなくデータ中の素性に重みをつける．そこではソース領域の訓練データのベクトル $\boldsymbol{x_s}$ を $(\boldsymbol{x_s}, \boldsymbol{x_s}, \mathbf{0})$ と連結した 3 倍の長さのベクトルに直し，ターゲット領域の訓練データのベクトル $\boldsymbol{x_t}$ を $(\mathbf{0}, \boldsymbol{x_t}, \boldsymbol{x_t})$ と連結した 3 倍の長さのベクトルに直す．ここで $\mathbf{0}$ は $\boldsymbol{x_s}$ や $\boldsymbol{x_t}$ と同じ次元数であり，しかもすべての次元の値が 0 であるようなベクトルである．

この 3 倍にしたベクトルを用いて，通常の分類問題として解く．この手法は非常に簡易でありながら，効果が高い手法として知られている．この拡張手法はソース領域とターゲット領域に共通している特徴が重なることで，結果として共通している特徴の重みがつくことで領域適応に効果が出ると考えられる．

また領域適応の問題を共変量シフト下の学習を用いて解決する研究としては，Jiang の研究 (Jiang and Zhai 2007) と齋木の研究 (齋木，高村，奥村 2008) がある．Jiang は確率密度比を手動で調整し，モデルにはロジステック回帰を用いている．また齋木は $P(\boldsymbol{x})$ を unigram でモデル化することで確率密度比を推定し，モデルには最大エントロピー法のモデルを用いている．ただしどちらの研究もタスクは WSD ではない．

また共変量シフト下では $P_S(c|\boldsymbol{x}) = P_T(c|\boldsymbol{x})$ を仮定するが，$P_S(\boldsymbol{x}|c) = P_T(\boldsymbol{x}|c)$ を仮定するアプローチもある．この場合，ベイズの定理から

$$\begin{aligned}\arg\max_{c\in C} P_T(c|\boldsymbol{x}) &= \arg\max_{c\in C} P_T(c)P_T(\boldsymbol{x}|c) \\ &= \arg\max_{c\in C} P_T(c)P_S(\boldsymbol{x}|c)\end{aligned}$$

となるので領域適応の問題は $P_T(c)$ の推定に帰着できる．実際，Chan らは $P_S(\boldsymbol{x}|c)$ と $P_T(\boldsymbol{x}|c)$ の違いの影響は非常に小さいと考え，$P_S(\boldsymbol{x}|c) = P_T(\boldsymbol{x}|c)$ を仮定し，$P_T(c)$ を EM アルゴリズムで推定することで WSD の領域適応を行っている (Chan and Ng 2005, 2006)．更に新納らは $P_S(\boldsymbol{x}|c) = P_T(\boldsymbol{x}|c)$ の仮定があったとしても，コーパスのスパース性から単純に $P_T(\boldsymbol{x}|c)$ を $P_S(\boldsymbol{x}|c)$ で置き換えることはできないと考え，$P_T(c)$ の推定の問題と $P_T(\boldsymbol{x}|c)$ の推定の問題を個別に対処することを提案している (新納，佐々木 2013)．

3 期待損失最小化からみた共変量シフト

対象単語 w の語義の集合を C，また w の用例 $\boldsymbol{x}$ 内の w の語義を c と識別したときの損失関数を $l(\boldsymbol{x}, c, d)$ で表す．d は w の語義を識別する分類器である．$P_T(\boldsymbol{x}, c)$ をターゲット領域上の分布とすれば，領域適応の問題における期待損失 L_0 は以下で表せる．

$$L_0 = \sum_{\boldsymbol{x},c} l(\boldsymbol{x}, c, d)P_T(\boldsymbol{x}, c)$$

また $P_S(\boldsymbol{x}, c)$ をソース領域上の分布とすると以下が成立する．

$$L_0 = \sum_{\boldsymbol{x},c} l(\boldsymbol{x}, c) \frac{P_T(\boldsymbol{x}, c)}{P_S(\boldsymbol{x}, c)} P_S(\boldsymbol{x}, c)$$

ここで共変量シフトの仮定から

$$\frac{P_T(\boldsymbol{x}, c)}{P_S(\boldsymbol{x}, c)} = \frac{P_T(\boldsymbol{x}) P_T(c|\boldsymbol{x})}{P_S(\boldsymbol{x}) P_S(c|\boldsymbol{x})} = \frac{P_T(\boldsymbol{x})}{P_S(\boldsymbol{x})}$$

となり，$r(\boldsymbol{x}) = P_T(\boldsymbol{x})/P_S(\boldsymbol{x})$ とおくと以下が成立する．

$$L_0 = \sum_{\boldsymbol{x},c} r(\boldsymbol{x}) l(\boldsymbol{x}, c, d) P_S(\boldsymbol{x}, c)$$

訓練データを $D = \{(\boldsymbol{x_i}, c_i)\}_{i=1}^N$ とし，$P_S(\boldsymbol{x}, c)$ を経験分布で近似すれば，

$$L_0 \approx \frac{1}{N} \sum_{i=1}^{N} r(\boldsymbol{x_i}) l(\boldsymbol{x_i}, c_i, d)$$

となるので，期待損失最小化の観点から考えると，共変量シフトの問題は以下の式 L_1 を最小にする d を求めればよいことがわかる．

$$L_1 = \sum_{i=1}^{N} r(\boldsymbol{x_i}) l(\boldsymbol{x_i}, c_i, d) \tag{1}$$

4 重み付き対数尤度の最大化

分類器 d として以下の事後確率最大化推定に基づく識別を考える．

$$d(\boldsymbol{x}) = \arg\max_c P_T(c|\boldsymbol{x})$$

また損失関数として対数損失 $-\log P_T(c|\boldsymbol{x})$ を用いれば，式 (1) は以下となる．

$$L_1 = -\sum_{i=1}^{N} r(\boldsymbol{x_i}) \log P_T(c|\boldsymbol{x_i})$$

つまり，分類問題の解決に $P_T(c|\boldsymbol{x}, \boldsymbol{\lambda})$ のモデルを導入するアプローチを取る場合，共変量シフト下での学習では，確率密度比を重みとした以下に示す重み付き対数尤度 $L(\boldsymbol{\lambda})$ を最大化するパラメータ $\boldsymbol{\lambda}$ を求める形となる．

$$L(\boldsymbol{\lambda}) = \sum_{i=1}^{N} r(\boldsymbol{x_i}) \log P(c_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) \tag{2}$$

ここではモデルとして以下の式で示される最大エントロピー法を用いる．

$$P_T(c|\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{Z(\boldsymbol{x},\boldsymbol{\lambda})}\exp\left(\sum_{j=1}^{M}\lambda_j f_j(\boldsymbol{x},c)\right) \tag{3}$$

$\boldsymbol{x} = (x_1, x_2, \cdots, x_M)$ が入力で c がクラスである．関数 $f_j(\boldsymbol{x},c)$ は素性関数であり，実質 $\boldsymbol{x}$ の真のクラスが c のときに x_j を返し，そうでないとき 0 を返す関数に設定される．$Z(\boldsymbol{x},\boldsymbol{\lambda})$ は正規化項であり，以下で表せる．

$$Z(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{c\in C}\exp\left(\sum_{j=1}^{M}\lambda_j f_j(\boldsymbol{x},c)\right) \tag{4}$$

そして $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_M)$ が素性に対応する重みパラメータとなる．

共変量シフト下ではない通常のケースでは，重みパラメータは最尤法から求める．つまり，訓練データ $D = \{(\boldsymbol{x_i}, c_i)\}_{i=1}^{N}$ とすると，以下の式 $F(\boldsymbol{\lambda})$ を最大にする $\boldsymbol{\lambda}$ を求める．

$$F(\boldsymbol{\lambda}) = \sum_{i=1}^{N}\log P(c_i|\boldsymbol{x_i})$$

これを各 λ_j で偏微分し極値問題に直すと以下が成立する．

$$\frac{\partial F(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_{i=1}^{N} f_j(\boldsymbol{x_i}, c_i) - \sum_{i=1}^{N}\sum_{c\in C} P_T(c|\boldsymbol{x_i},\boldsymbol{\lambda}) f_j(\boldsymbol{x_i}, c) = 0$$

これを勾配法などで解くことにより $\boldsymbol{\lambda}$ が求まる．

共変量シフト下の学習では式 (2) の $L(\boldsymbol{\lambda})$ を最大にする $\boldsymbol{\lambda}$ を求める．上記と全く同じ手順で，

$$\frac{\partial L(\boldsymbol{\lambda})}{\partial \lambda_j} = \sum_{i=1}^{N} r(\boldsymbol{x_i}) f_j(\boldsymbol{x_i}, c_i) - \sum_{i=1}^{N}\sum_{c\in C} P(c|\boldsymbol{x_i},\boldsymbol{\lambda}) r(\boldsymbol{x_i}) f_j(\boldsymbol{x_i}, c) = 0$$

が得られる．これを勾配法などで解くことにより $\boldsymbol{\lambda}$ が求まる．

今，事例 $\boldsymbol{x_i}$ の頻度を h_i とすると，尤度は以下となる．

$$\prod_{i=1}^{N} P(c_i|\boldsymbol{x_i})^{h_i}$$

対数を取れば以下が得られる．

$$\sum_{i=1}^{N} h_i \log P(c_i|\boldsymbol{x_i})$$

この式は重み付き対数尤度の式 (2) と同じ形なので，実際に $\boldsymbol{\lambda}$ を求めるためには，事例 $\boldsymbol{x_i}$ の

頻度 h_i を $r(\boldsymbol{x_i})$ と考えて，最大エントロピー法のツールなどを用いればよい[3].

5 確率密度比の算出

共変量シフト下の学習では確率密度比の算出が鍵である．直接的には $P_S(\boldsymbol{x})$ と $P_T(\boldsymbol{x})$ を推定し，その比を取ればよいが，$P_S(\boldsymbol{x})$ や $P_T(\boldsymbol{x})$ を正確に推定することは困難であり，その比をとれば更に誤差が大きくなると予想できる．そのため確率密度比を直接モデル化して求める手法が活発に研究されている (杉山 2010).

ただし本稿では簡易な手法を利用して確率密度比を算出することにした．本稿の目的はこのような簡易な手法による確率密度比の算出法であっても，WSD の領域適応の有力な解法になることを示すことである.

対象単語 w の用例 $\boldsymbol{x}$ の素性リストを $\{f_1, f_2, \cdots, f_n\}$ とする．求めるのは領域 $R \in \{S, T\}$ 上の $\boldsymbol{x}$ の分布 $P_R(\boldsymbol{x})$ である．ここでは Naive Bayes で使われるモデルを用いて算出する．Naive Bayes のモデルでは以下を仮定する.

$$P_R(\boldsymbol{x}) = \prod_{i=1}^{n} P_R(f_i)$$

領域 R のコーパス内の w の全ての用例について素性リストを作成しておく．ここで用例の数を $N(R)$ とおく．また $N(R)$ 個の用例の中で，素性 f が現れた用例数を $n(R, f)$ とおく．MAP 推定でスムージングを行い，$P_R(f)$ を以下で定義する (高村 2010).

$$P_R(f) = \frac{n(R, f) + 1}{N(R) + 2}$$

以上より，ソース領域 S の用例 $\boldsymbol{x}$ に対して，確率密度比 $r(\boldsymbol{x}) = \frac{P_T(\boldsymbol{x})}{P_S(\boldsymbol{x})}$ が計算できる．ターゲット領域 T の用例 $\boldsymbol{x}$ に対しては $r(\boldsymbol{x}) = 1$ とする．また $r_x < 0.01$ となる用例 $\boldsymbol{x}$ は訓練データから削除した[4].

6 提案手法

「関連手法」の節で素性空間拡張法を紹介した．素性空間拡張法はデータの表現を領域適応で効果が出るように拡張する手法である．そして拡張されたデータに対しては任意の学習手法

[3] ただし利用できるツールは頻度を実数値として与えられるものでなくてはならない．事例の重みを頻度の拡張として実装したツールであるともいえる．本稿で用いた機械学習ツール Classias (Okazaki 2009) はこの条件を満たすため利用可能である.

[4] この削除は処理の効率化のために行っている．また本稿の実験では削除しない場合よりもわずかによい結果となっていた.

が利用できる．つまり拡張されたデータに対して，共変量シフト下の学習も可能である．本稿では，素性空間拡張法により拡張されたデータに対して，4 章で説明した共変量シフト下の学習を行うことを提案手法する．

具体的に示す．素性空間拡張法により，ソース領域の訓練データ $\boldsymbol{x_s}$ は $\boldsymbol{u_s} = (\boldsymbol{x_s}, \boldsymbol{x_s}, \boldsymbol{0})$ という 3 倍の長さのベクトルに拡張され，ターゲット領域の訓練データ $\boldsymbol{x_t}$ は $\boldsymbol{u_t} = (\boldsymbol{0}, \boldsymbol{x_t}, \boldsymbol{x_t})$ という 3 倍の長さのベクトルに拡張される．ここで $\boldsymbol{u_s}$ に対しては確率密度比 $r(\boldsymbol{x_s}) = P_T(\boldsymbol{x_s})/P_S(\boldsymbol{x_s})$ の重みをつけ，$\boldsymbol{u_t}$ に対しては重み 1 をつける．また $P_T(c|\boldsymbol{u})$ のモデルに最大エントロピー法を用い，重み付き対数尤度を最大化するパラメータを求めることで，$P(c|\boldsymbol{u})$ を推定する．

上記の重み付き対数尤度の式（目的関数）を示しておく．今，ソース領域の訓練データを $D_s = \{(\boldsymbol{x_s^{(i)}}, c_s^{(i)})\}_{i=1}^n$，ターゲット領域の訓練データを $D_t = \{(\boldsymbol{x_t^{(i)}}, c_t^{(i)})\}_{i=1}^m$ とおく．また $\boldsymbol{x_s^{(i)}}$ と $\boldsymbol{x_t^{(i)}}$ を素性空間拡張法により拡張したデータをそれぞれ $\boldsymbol{u_s^{(i)}}$ と $\boldsymbol{u_t^{(i)}}$ とおく．ここで $\boldsymbol{x_s^{(i)}}$ と $\boldsymbol{x_t^{(i)}}$ は M 次元，$\boldsymbol{u_s^{(i)}}$ と $\boldsymbol{u_t^{(i)}}$ は $3M$ 次元のベクトルであることに注意する．提案手法の重み付き対数尤度の式は以下となる．

$$L(\boldsymbol{\lambda}) = \sum_{i=1}^{n} r(\boldsymbol{x_s^{(i)}}) \log P(c_s^{(i)}|\boldsymbol{u_s^{(i)}}, \boldsymbol{\lambda}) + \sum_{i=1}^{m} \log P(c_t^{(i)}|\boldsymbol{u_t^{(i)}}, \boldsymbol{\lambda})$$

$$P(c|\boldsymbol{u}, \boldsymbol{\lambda}) = \frac{1}{Z(\boldsymbol{u}, \boldsymbol{\lambda})} \exp\left(\sum_{j=1}^{3M} \lambda_j f_j(\boldsymbol{u}, c)\right)$$

$$Z(\boldsymbol{u}, \boldsymbol{\lambda}) = \sum_{c \in C} \exp\left(\sum_{j=1}^{3M} \lambda_j f_j(\boldsymbol{u}, c)\right)$$

7 実験

BCCWJ コーパスの PB（書籍），OC （Yahoo! 知恵袋）及び PN （新聞）を異なった領域として実験を行う．SemEval-2 の日本語 WSD タスク (Okumura et al. 2010) ではこれら領域のコーパスの一部に語義タグを付けたデータを公開しており，そのデータを利用する．この 3 つの領域からある程度頻度のある多義語 16 単語を WSD の対象単語とする．これら単語と辞書上での語義数及び各コーパスでの頻度と語義数を表 1 に示す[5]．領域適応の方向としては OC → PB，PB → PN，PN → OC，OC → PN，PN → PB，PB → OC の計 6 通りの方向が存在する．

本稿で利用した素性は以下の 8 種類である．(e0) w の表記，(e1) w の品詞，(e2) w_{-1} の表記，(e3) w_{-1} の品詞，(e4) w_1 の表記，(e5) w_1 の品詞，(e6) w の前後 3 単語までの自立語の表

[5] 語義は岩波国語辞書がもとになっている．そこでの中分類までを対象にした．また「入る」は辞書上の語義が 3 つだが，OC や PB では 4 つの語義がある．これは SemEval-2 の日本語 WSD タスクでは新語義のタグも許しているからである．

表 1 対象単語

単語	辞書上の語義数	OC での頻度	OC での語義数	PB での頻度	PB での語義数	PN での頻度	PN での語義数
言う	3	666	2	1114	2	363	2
入れる	3	73	2	56	3	32	2
書く	2	99	2	62	2	27	2
聞く	3	124	2	123	2	52	2
子供	2	77	2	93	2	29	2
時間	4	53	2	74	2	59	2
自分	2	128	2	308	2	71	2
出る	3	131	3	152	3	89	3
取る	8	61	7	81	7	43	7
場合	2	126	2	137	2	73	2
入る	3	68	4	118	4	65	3
前	3	105	3	160	2	106	4
見る	6	262	5	273	6	87	3
持つ	4	62	4	153	3	59	3
やる	5	117	3	156	4	27	2
ゆく	2	219	2	133	2	27	2
平均	3.44	148.19	2.94	199.56	3.00	75.56	2.69

記，(e7) e6 の分類語彙表の番号の 4 桁と 5 桁．なお対象単語の直前の単語を w_{-1}，直後の単語を w_1 としている．

単語 w_i についてソース領域 S からターゲット領域 T への領域適応の実験について説明する．まずターゲット領域 T のラベル付きデータをランダムに 15 個取り出し，残りを評価データとする．つまり利用できる訓練データはソース領域 S のラベル付きデータとターゲット領域 T からランダムに取り出した 15 個のラベル付きデータとなる．この訓練データを用いて手法 A により分類器を作成し，先の評価データの語義識別の正解率 $P_{i,k}$ を測る．この実験を 5 回行い $P_{i,1}, P_{i,2}, \cdots, P_{i,5}$ を得る．それらの平均 P_i を「単語 w_i の S から T への領域適応における手法 A の平均正解率」とする．上記の単語 w_i を 16 種類の各対象単語 $w_1, w_2, \cdots, w_{16}$ に変えることで，16 個の平均正解率 $P_1, P_2, \cdots, P_{16}$ が得られる．それらの平均 P を「S から T への領域適応における手法 A の平均正解率」とする（図 1 参照）．

上記の手法 A としては，以下の 6 種類を試す．(1) ソース領域のラベル付きデータのみを用いる手法（ターゲット領域の 15 個のラベル付きデータの重みを 0 とする手法）(S-Only)，(2) ターゲット領域からランダムに取り出した 15 個のラベル付きデータのみを用いる手法（ソース領域のラベル付きデータの重みを 0 とする手法）(T-Only)，(3) ソース領域のラベル付きデータとターゲット領域の 15 個のラベル付きデータを用いる手法 (S+T)，(4) Daumé の手法 (Daumé)，(5)

本稿で示した簡易手法により算出した確率密度比を用いた共変量シフトによる手法 (Cov-Shift), (6) 素性空間拡張法から得られた訓練データに対して，本稿で示した簡易手法により算出した確率密度比を用いた共変量シフトによる手法（提案手法）の計 6 種類である．またすべての手法において学習アルゴリズムとしては最大エントロピー法を用いた．またその実行にはツールの Classias を用いた (Okazaki 2009).

S から T への領域適応における各手法の平均正解率を表 2 に示す．Daumé と Cov-Shift を比

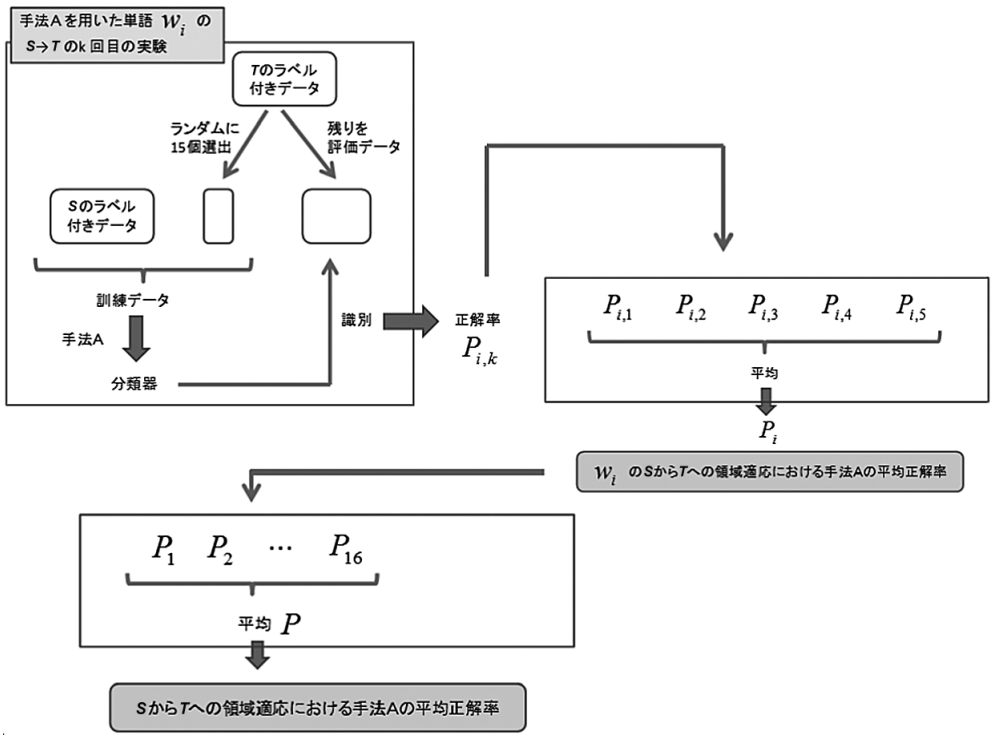

図 1 手法の評価値（平均正解率）の算出

表 2 各手法の平均正解率

領域適応	S-Only	T-only	S+T	Daumé	Cov-Shift	提案手法
OC → PB	0.7137	0.7559	0.7511	0.7599	0.7567	**0.7621**
PB → PN	0.7678	0.7206	0.7801	**0.7859**	0.7847	0.7805
PN → OC	0.6926	0.7716	0.7630	0.7704	0.7781	**0.7856**
OC → PN	0.6829	0.7300	0.7324	0.7368	0.7373	**0.7369**
PN → PB	0.7543	0.7561	**0.7863**	0.7850	0.7824	0.7841
PB → OC	0.6988	0.7766	0.7533	0.7772	0.7642	**0.7823**
平均	0.7184	0.7518	0.7611	0.7692	0.7672	**0.7719**

較すると Daumé の方がわずかに高い正解率を示している．この点は考察で議論する．ただし提案手法は Daumé よりも高い正解率であり，共変量シフトによる解法の効果が確認できる．

8 考察

8.1 負の転移の有無

WSD の領域適応では，対象単語毎に領域適応の問題が生じている．実験では領域の組み合わせで 6 通り，対象単語が 16 単語あるので，合計 96 ($= 6 \times 16$) 通りの領域適応の問題を扱ったことになる．ここでは各領域適応の問題に対して負の転移が生じているかどうかを調べ，それぞれのケースに分けて，各手法の正解率を調べた．

まず負の転移が生じているかどうかの判定には，先の実験でより得られた `T-Only`，`S-Only` 及び `S+T` の正解率を利用する．もしも正解率で以下の関係が成立しているなら，負の転移が生じていないと考えられる．

```
T-Only, S-Only  <  S+T
```

結果を表 3 に示す．チェックがつけられた箇所が負の転移が生じていない領域適応の問題である．96 種類の領域適応の問題の中で 44 種類において負の転移が生じていない．

次に負の転移が生じているかいないかのケースに分けて，各手法の平均正解率を調べた．結果を表 4 に示す．

表 3 負の転移が生じていない領域適応

単語	OC → PB	PB → PN	PN → OC	OC → PN	PN → PB	PB → OC
言う		✓	✓	✓	✓	
入れる		✓	✓	✓	✓	✓
書く	✓			✓	✓	
聞く	✓					
子供			✓		✓	
時間	✓		✓		✓	
自分	✓	✓				
出る				✓		✓
取る			✓		✓	✓
場合		✓	✓		✓	✓
入る		✓	✓		✓	✓
前		✓				
見る	✓					
持つ	✓	✓				✓
やる		✓		✓		✓
ゆく		✓		✓	✓	

表 4 負の転移と各手法の平均正解率

負の転移	S-Only	T-Only	S+T	Daumé	Cov-Shift	提案手法
無し (44)	0.7542	0.6985	0.8020	0.8755	0.8760	**0.8846**
有り (52)	0.6880	**0.7969**	0.7264	0.6792	0.6752	0.6765
平均	0.7184	0.7518	0.7611	0.7692	0.7672	**0.7719**

表 4 において領域適応に対処する 3 手法（Daumé，Cov-Shift，提案手法）を見ると，提案手法は負の転移の有無に関わらず Cov-Shift よりも高い正解率であり，提案手法は Cov-Shift の改良になっていることがわかる．更に負の転移が生じていないケースでは Cov-Shift は Daumé よりも正解率が高く，このケースでは素性に重みをつけるよりも事例に重みをつける方が効果があることがわかる．ただし負の転移が生じるケースでは，提案手法は Daumé よりも正解率が若干低い．つまり提案手法を Daumé の手法の改良と見た場合，負の転移が生じるケースでは正解率の低下を抑え，その代わりに負の転移が生じないケースで正解率を高めることで，全体的な正解率を改善する手法と見なせる．

また領域適応に対処しない 3 手法 (S-Only, T-Only, S+T) も含めて比較すると，負の転移が生じるケースでは領域適応に対処する 3 手法（Daumé，Cov-Shift，提案手法）の正解率はかなり悪い．つまり WSD の領域適応では負の転移を検出することで大きな改善が期待できる．共変量シフト下の学習では，負の転移が生じているケースに対しては，ソース領域のデータに 0 に近い重みを与えられればよいはずである．より正確な確率密度比の推定法を利用することで，このような重み付けが可能だと考える．この点は今後の課題である．

8.2 確率密度比の調整

確率密度比を精度良く推定することは困難な問題である．そのために求まった確率密度比を調整することも行われている．杉山は確率密度比 r に p $(0 < p < 1)$ 乗した r^p を重みにすることを提案している (杉山 2006)．また Yamada は relative density ratio として確率密度比を以下の形で求めることを提案してる (Yamada, Suzuki, Kanamori, Hachiya, and Sugiyama 2011)．

$$\frac{P_T(\boldsymbol{x})}{\alpha P_T(\boldsymbol{x}) + (1-\alpha) P_S(\boldsymbol{x})}$$

ここでは $r^{0.5}$ の重みと $\alpha = 0.5$ の relative density ratio を試した．結果を表 5 に示す．表 5 における提案手法と Cov-Shift は表 2 における提案手法と Cov-Shift と同じものである．$r^{0.5}$ が Cov-Shift の重み r を 0.5 乗したものであり，RDR が $\alpha = 0.5$ の relative density ratio である．

表 5 をみると，$r^{0.5}$ や relative density ratio の調整は一部有効な問題もあったが，全体として見ると，効果はあまりない．これも本来の確率密度値 $P_S(\boldsymbol{x})$ や $P_T(\boldsymbol{x})$ の推定が簡易すぎるために生じていると考える．

確率密度比を確率統計的により精緻に求めていくことは重要である．ただし確率密度比は事例の重み，つまり事例の重要度を意味している．事例の重要度という自然言語処理的な観点から WSD の領域適応に特化した重みの設定も可能である．

8.3 SVM の利用

本稿では学習アルゴリズムとして最大エントロピー法を用いた．共変量シフトの解法として，重み付き対数尤度を最大化する形では，$P_T(c|\boldsymbol{x})$ をモデル化するアプローチに限られる．しかし共変量シフト下の学習では確率密度比を重みにして期待損失を最小化すれば良いので，損失関数ベースの学習手法が利用できる．例えばヒンジ損失関数に密度比で重みづけすることで共変量シフト下の学習に SVM を利用できる (Sugiyama and Kawanabe 2011)．ただし SVM 自体の実装が容易ではないために簡単に試すことはできない．

ここでは共変量シフト下の学習に SVM を用いるのではなく，素性空間拡張法により拡張されたデータに対して，SVM を利用してみる．実行にはツールの libsvm[6]を用いた．またそこで利用したカーネルは線形カーネルである．実験結果を表 6 に示す．

提案手法が本稿での提案手法での平均正解率であり，D3-ME が素性空間拡張法と最大エント

表 5 確率密度比の調整による平均正解率

領域適応	提案手法	Cov-shft	$r^{0.5}$	RDR
OC → PB	0.7621	0.7567	**0.7652**	0.7517
PB → PN	0.7805	**0.7847**	0.7750	0.7815
PN → OC	**0.7856**	0.7781	0.7742	0.7643
OC → PN	0.7369	**0.7373**	0.7331	0.7360
PN → PB	0.7841	0.7824	0.7654	**0.7854**
PB → OC	0.7823	0.7642	**0.7940**	0.7530
平均	**0.7719**	0.7672	0.7678	0.7620

表 6 SVM による平均正解率

領域適応	提案手法	D3-ME	D3-SVM
OC → PB	0.7621	0.7599	**0.7678**
PB → PN	0.7805	**0.7859**	0.7844
PN → OC	**0.7856**	0.7704	0.7736
OC → PN	0.7369	0.7368	**0.7373**
PN → PB	0.7841	0.7850	**0.7905**
PB → OC	**0.7823**	0.7772	0.7749
平均	**0.7719**	0.7692	0.7714

[6] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

ロピー法を利用した場合の平均正解率である．つまり提案手法と D3-ME は，表 2 での提案手法と Daumé に対応する．そして D3-SVM が素性空間拡張法と SVM を利用した場合の平均正解率である．提案手法は D3-SVM よりもわずかに高い正解率となっているが，その差は小さく識別能力については，提案手法と D3-SVM は同程度と言える．また D3-SVM は D3-ME よりも正解率が高い．つまり最大エントロピー法ではなく，SVM を利用する方が正解率が高くなると予想できる．このことから共変量シフト下の学習に SVM を利用すれば，改善が可能であると考えられる．これは今後の課題である．

8.4 教師なし手法への適用

共変量シフト下での学習では訓練データの中にターゲット領域のデータが含まれる必要はない．ターゲット領域の訓練データを含めなければ，教師なし領域適応手法となるはずである．この点を確認した実験を行った．実験結果を表 7 に示す．表の S-Only の列はソース領域の訓練データだけで学習した結果である．これは表 2 の S-Only に対応する．W-S-Only はソース領域の訓練データのみを使った共変量シフト下での学習手法である．また参考までに提案手法の結果も記している．

確率密度比を用いる W-S-Only ではソース領域のデータへの重みが小さくなりがちである．ここでの実験では重みが 0.01 未満の場合はそのデータを省いて学習させている．そのために W-S-Only では極端にラベル付きデータが減少するケースがあった．結果として精度が低くなってしまったと考えられる．また多くの単語で正解率の低下が起こっていた．この原因としては，重みのあるデータの欠如だと考える．例えば，語義 c_1 のデータ x_1 の重みが 0.01，語義 c_2 のデータ x_2 の重みが 0.02 である場合，どちらの重みも「小さく」，その差はほぼ等しいと見なして $P(c_1) = P(c_2) = 0.5$ と考えるのが妥当であるが，「小さい」という点を考えないと $P(c_1) = 1/3$, $P(c_2) = 2/3$ となってしまう．「小さい」という点を考えるためには比較となるある程度「大きな」データが必要である．例えば，上記の設定の上で語義 c_1 のデータ x_3 の重みが 1 などとい

表 7 重み付き教師なし学習による平均正解率

領域適応	提案手法（教師あり）	S-Only（教師なし）	W-S-Only（教師なし）
OC → PB	0.7622	**0.7137**	0.7129
PB → PN	0.7861	0.7678	**0.7725**
PN → OC	0.7702	**0.6926**	0.6718
OC → PN	0.7447	0.6829	**0.6846**
PN → PB	0.7874	**0.7543**	0.6667
PB → OC	0.7750	**0.6988**	0.6799
平均	0.7716	**0.7184**	0.6981

うデータが存在すれば，$P(c_1) = 101/103$, $P(c_2) = 2/103$ となり，これは妥当である．つまり重みが低いデータが多数を占めるような場合，信頼性のある推定が行えない．ある程度，重みのあるデータが必要だと思われる．このため共変量シフト下での学習を教師なしの枠組みに単純に利用することは難しい．教師なしの枠組みへの利用方法の検討は今後の課題である．

9 おわりに

本稿では WSD の領域適応の問題が共変量シフトの問題と見なせることを示した．そして，共変量シフトの標準的な解法である確率密度比を重みにしたパラメータ学習により，WSD の領域適応の解決が図れることを示した．また素性空間拡張法により拡張されたデータに対して，共変量シフトの解法を行う手法を提案した．

BCCWJ コーパスの 3 つ領域 OC （Yahoo! 知恵袋），PB （書籍）及び PN （新聞）を選び，SemEval-2 の日本語 WSD タスクのデータを利用して，上記領域にある程度の頻度がある多義語 16 単語を対象に，WSD の領域適応の実験を行った．実験の結果，提案手法は Daumé の手法と同等以上の正解率を出した．

共変量シフトの解法では確率密度比の算出が鍵となるが，ここでは Naive Bayes で利用されるモデルを利用した簡易な算出法を試みた．このような簡易な算出法であっても WSD の領域適応に共変量シフトの解法を利用する効果が高いことが示された．

より正確な確率密度比の推定法を利用したり，最大エントロピー法に代えて SVM を利用するなどの工夫で更なる改善が可能である．また教師なし領域適応へも応用可能である．WSD の領域適応に共変量シフトの解法を利用することは有望であると考えられる．

謝 辞

Classias の作者である岡崎直観氏に，Classias の事例の重み付け方法について教えていただきました．また本稿の査読者殿には有益なコメントいただきました．感謝いたします．

参考文献

Chan, Y. S. and Ng, H. T. (2005). “Word Sense Disambiguation with Distribution Estimation.” In *Proceedings of IJCAI-2005*, pp. 1010–1015.

Chan, Y. S. and Ng, H. T. (2006). “Estimating class priors in domain adaptation for word sense disambiguation.” In *Proceedings of COLING-ACL-2006*, pp. 89–96.

Daumé, H. I. (2007). “Frustratingly Easy Domain Adaptation.” In *Proceedings of ACL-2007*, pp. 256–263.

張本佳子，宮尾祐介，辻井潤一 (2010). 構文解析の分野適応における精度低下要因の分析及び分野間距離の測定手法. 言語処理学会第 16 回年次大会, pp. 27–30.

Jiang, J. and Zhai, C. (2007). “Instance weighting for domain adaptation in NLP.” In *Proceedings of ACL-2007*, pp. 264–271.

神嶌敏弘 (2010). 転移学習. 人工知能学会誌, **25** (4), pp. 572–580.

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, **19** (3), pp. 143–166.

Komiya, K. and Okumura, M. (2011). “Automatic Determination of a Domain Adaptation Method for Word Sense Disambiguation using Decision Tree Learning.” In *Proceedings of IJCNLP-2011*, pp. 1107–1115.

Komiya, K. and Okumura, M. (2012). “Automatic Domain Adaptation for Word Sense Disambiguation Based on Comparison of Multiple Classifiers.” In *Proceedings of PACLIC-2012*, pp. 75–85.

Maekawa, K. (2007). “Design of a Balanced Corpus of Contemporary Written Japanese.” In *Symposium on Large-Scale Knowledge Resources (LKR2007)*, pp. 55–58.

Okazaki, N. (2009). “Classias: a collection of machine-learning algorithms for classification.” `http://www.chokkan.org/software/classias/`.

Okumura, M., Shirai, K., Komiya, K., and Yokono, H. (2010). “SemEval-2010 Task: Japanese WSD.” In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 69–74.

Plank, B. and van Noord, G. (2011). “Effective measures of domain similarity for parsing.” In *Proceedings of ACL-2011*, pp. 1566–1576.

Ponomareva, N. and Thelwall, M. (2012). “Which resource is best for cross-domain sentiment analysis?” In *Proceedings of CICLing-2012*, pp. 488–499.

Remus, R. (2012). “Domain Adaptation Using Domain Similarity- and Domain Complexity-based Instance Selection for Cross-domain Sentiment Analysis.” In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW 2012) Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE)*, pp. 717–723.

Rosenstein, M. T., Marx, Z., Kaelbling, L. P., and Dietterich, T. G. (2005). “To transfer or not to transfer.” In *Proceedings of the NIPS 2005 Workshop on Inductive Transfer: 10 Years Later*.

齋木陽介，高村大也，奥村学 (2008). 文の感情極性判定における事例重み付けによるドメイン

適応. 情報処理学会第 184 回自然言語処理研究会, pp. 61–67.

Shimodaira, H. (2000). "Improving predictive inference under covariate shift by weighting the log-likelihood function." *Journal of statistical planning and inference*, **90** (2), pp. 227–244.

新納浩幸, 佐々木稔 (2013). k 近傍法とトピックモデルを利用した語義曖昧性解消の領域適応. 自然言語処理, **20** (5), pp. 707–726.

Sogaard, A. (2013). *Semi-Supervised Learning and Domain Adaptation in Natural Language Processing*. Morgan & Claypool.

杉山将 (2006). 共変量シフト下での教師付き学習. 日本神経回路学会誌, **13** (3), pp. 111–118.

杉山将 (2010). 密度比に基づく機械学習の新たなアプローチ. 統計数理, **58** (2), pp. 141–155.

Sugiyama, M. and Kawanabe, M. (2011). *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. MIT Press.

高村大也 (2010). 言語処理のための機械学習入門. コロナ社.

Van Asch, V. and Daelemans, W. (2010). "Using domain similarity for performance estimation." In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pp. 31–36.

Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., and Sugiyama, M. (2011). "Relative density-ratio estimation for robust distribution comparison." *Neural Computation*, **25** (5), pp. 1370–1370.

略歴

新納　浩幸：1985 年東京工業大学理学部情報科学科卒業. 1987 年同大学大学院理工学研究科情報科学専攻修士課程修了. 同年富士ゼロックス, 翌年松下電器を経て, 1993 年 4 月茨城大学工学部システム工学科助手. 1997 年 10 月同学科講師, 2001 年 4 月同学科助教授, 現在, 茨城大学工学部情報工学科准教授. 博士（工学）. 機械学習や統計的手法による自然言語処理の研究に従事. 言語処理学会, 情報処理学会, 人工知能学会 各会員.

佐々木　稔：1996 年徳島大学工学部知能情報工学科卒業. 2001 年同大学大学院博士後期課程修了. 博士（工学）. 2001 年 12 月茨城大学工学部情報工学科助手. 現在, 茨城大学工学部情報工学科講師. 機械学習や統計的手法による情報検索, 自然言語処理等に関する研究に従事. 言語処理学会, 情報処理学会 各会員.

（2013 年 9 月 16 日　受付）
（2013 年 11 月 5 日　再受付）
（2013 年 12 月 10 日　再々受付）
（2013 年 12 月 27 日　採録）