uLSIF による重み付き学習を利用した語義曖昧性解消の領域適応

菊池 裕紀 (茨城大学大学院 情報工学専攻) [1]
新納 浩幸 (茨城大学 工学部 情報工学科) [2]

Domain Adaptation for Word Sense Disambiguation by Weighted Learning using uLSIF

Hironori Kikuchi (Ibaraki University, Department of Computer and Information Sciences)
Hiroyuki Shinnou (Ibaraki University, Department of Computer and Information Sciences)

1 はじめに

本論文では，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を共変量シフトの問題であるとみなし，その問題の解法である確率密度比を重みとしたパラメータ学習により WSD の領域適応の問題解決を図る．確率密度比の算出には，Naive Bayes のモデルを利用して算出したものと uLSIF によって算出したものを用いて比較実験を行う．

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 A 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タスクの解決を図っていく．このとき，実際のタスクとなるデータはコーパス A とは領域が異なるコーパス B 内のものであることがしばしば起こる．この場合，コーパス A（ソースドメイン）から学習された分類器では，コーパス B（ターゲットドメイン）のデータを精度よく解析できない問題が生じる．これが領域適応の問題であり，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 m の語義 $c \in C$ を識別する問題である．事後確率最大化に基づければ，$\arg\max_{c \in C} P(c|\boldsymbol{x})$ を解く問題といえる．例えば，単語 $m =$「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．通常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．しかし，WSD の領域適応の問題は，前述したように，学習もとのソースドメインのコーパスと，分類器の適用先であるターゲットドメインのコーパスが異なる問題である．領域適応ではソースドメイン S から S 上の確率密度 $P_s(c|\boldsymbol{x})$ 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，ターゲットドメイン T 上の確率密度 $P_t(c|\boldsymbol{x})$ 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，この文がどんな領域のコーパスに現れても変化するとは考えられない．つまり $P(c|\boldsymbol{x})$ は領域に依存していないと考えられる．つまり $P_s(c|\boldsymbol{x}) = P_t(c|\boldsymbol{x})$ が成立している．今 $P_s(c|\boldsymbol{x})$ は学習できるので，$P_s(c|\boldsymbol{x}) = P_t(c|\boldsymbol{x})$ 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．ただしソースドメイン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) 杉山将 (2006)) である．

訓練データを $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ とする．共変量シフトでは $P_t(y|\boldsymbol{x})$ に確率モデル $P(y|\boldsymbol{x};\theta)$ を設定する．次に確率密度比 $w(\boldsymbol{x}_i) = P_t(\boldsymbol{x}_i)/P_s(\boldsymbol{x}_i)$ を重みにした以下の対数尤度を最大にする θ を求めることで，$P_t(y|\boldsymbol{x})$ を構築する．

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

実験では BCCWJ コーパス Maekawa (2007) の 3 つ領域 OC (Yahoo! 知恵袋) PB (書籍) と PN（新聞）から共に頻度が 50 以上の多義語 16 単語を対象にして，WSD の領域適応の実験を行い，前

[2] 13nm705g@hcs.ibaraki.ac.jp
[2] shinnou@mx.ibaraki.ac.jp

述した共変量シフトの問題として WSD の領域適応の問題を解いた．$P_t(y|\boldsymbol{x})$ の確率モデルとしては Naive Bayes（NB） のモデルを用いた．

2　関連研究

自然言語処理に対する領域適応は，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．そのため，その研究は多岐に渡っている．その中でも現在注目しているのは，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同値であるとするものである．この場合，WSD の領域適応の問題は，共変量シフトの解決策である確率密度比を重みとしたパラメータ学習により解消できると考えられる．実際，この解決手法は古くから扱われている (Shimodaira (2000))(杉山将 (2006)).

教師付き領域適応手法では標準手法とも言える Daumé(Daumé III, Hal (2007)) の手法も重み付けの手法である．ただしこれは事例への重みではなく，素性への重みである．そこではソースドメインの訓練データのベクトル x_s を $(x_s, x_s, 0)$ と連結した 3 倍の長さのベクトルになるよう形成し，同様にターゲットドメインの訓練データのベクトル x_t を 3 倍の長さのベクトル $(0, x_t, x_t)$ と形成する．このベクトルを用いて，通常の分類問題として解いていく．この手法は非常に簡単でありながら，効果が高い手法として知られている．この手法は，ソースドメインとターゲットドメインに共通している特徴の重みが付くことで領域適応に効果が出ていると考えられる．

また，領域適応の問題を共変量シフト下の学習とし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 (Yosuke Saiki and Okumura (2008)) が挙げられる．Jiang は密度比を手動で調整し，モデルにはロジスティック回帰を用いている．また齋木は $P(x)$ を unigram でモデル化することで密度比を推定し，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に WSD を扱っていない．更に確率密度比 $w(\boldsymbol{x})$ の算出には結局 $P_s(\boldsymbol{x})$ と $P_t(\boldsymbol{x})$ をそれぞれ求めてその比を取る手法になっており，本論文利用した uLSIF のように直接 $w(\boldsymbol{x})$ を推定する手法は用いられていない．

3　確率密度比

確率密度比 $w(\boldsymbol{x}) = P_t(\boldsymbol{x})/P_s(\boldsymbol{x})$ の算出には大きく 2 つの方法がある． 1 つは $P_t(\boldsymbol{x})$ と $P_s(\boldsymbol{x})$ を各々求め，その比をとる方法であり，もう 1 つは直接 $w(\boldsymbol{x})$ をモデル化して推定する方法である．ここでは前者として論文 (新納浩幸・佐々木稔 (2014)) で提案された簡易な手法を利用する．また後者として uLSIF を利用する．

3.1　NaiveBayes による確率密度比の推定

対象単語 m 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 $P_R(\boldsymbol{x})$ である．ここでは以下で仮定される NB のモデルを用いて算出する．

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

領域 R のコーパス内の m 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．高村大也 (2010)

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

以上で求められた $P_s(\boldsymbol{x})$ と $P_t(\boldsymbol{x})$ の比を求めることで $\widehat{w}(\boldsymbol{x})$ を求める．

$$\widehat{w}(\boldsymbol{x}) = \frac{P_t(\boldsymbol{x})}{P_s(\boldsymbol{x})}$$

3.2 uLSIF による確率密度比の算出

uLSIF は Kanamori ら (Kanamori et al. (2009)) により考案された手法である．$w(\boldsymbol{x})$ を線形モデルで直接推定することができる．線型方程式を解くことによってモデルのパラメータを推定できるため高速な算出が可能となっている．

$w(\boldsymbol{x})$ を以下のモデルで定式化する．

$$\widehat{w} = \sum_{l=1}^{b} \boldsymbol{\alpha_l \varphi_l}(\boldsymbol{x}) \tag{1}$$

ここで，$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_b)^T$ はサンプルデータから学習されるパラメータを，$\varphi_l(\boldsymbol{x})$ は，以下のような基底関数を表す.

$$\varphi_l(x) >= 0 \ \forall x \in D \ \forall l = 1, 2, \cdots, b \tag{2}$$

b は $R \in \{S, T\}$ のデータ数により決定する．また，$\varphi_l(\boldsymbol{x})$ には以下のガウスカーネルを用いることとする．

$$K_\sigma(\boldsymbol{x}, \boldsymbol{x'}) = \exp\bigl(-\frac{\|\boldsymbol{x} - \boldsymbol{x'}\|^2}{2\sigma^2}\bigr) \tag{3}$$

よって，式 (1) は，以下で表すことができる．

$$\widehat{w}(x) = \sum_{l=1}^{n_T} \boldsymbol{\alpha}_l K_\sigma(x, x_l^T) \tag{4}$$

次に，パラメータ $\boldsymbol{\alpha}_l$ の求め方を定義する．パラメータ $\boldsymbol{\alpha}_l$ は，次の J_0 を最小にするような二乗誤差を求めればよい.

$$\begin{aligned} J_0(\boldsymbol{\alpha}) &= \frac{1}{2}\int (\widehat{w}(x) - w(x))^2 p_S(x)dx \\ &= \frac{1}{2}\int \widehat{w}(x)^2 P_s(\boldsymbol{x})dx - \int \widehat{w}(x)w(x)P_s(\boldsymbol{x})dx + \frac{1}{2}\int w(x)^2 P_s(\boldsymbol{x})dx \\ &= \frac{1}{2}\int \widehat{w}(x)^2 P_s(\boldsymbol{x})dx - \int \widehat{w}(x)P_t(\boldsymbol{x})dx + \frac{1}{2}\int w(x)^2 P_s(\boldsymbol{x})dx \end{aligned} \tag{5}$$

なお，式 (5) の最終項は一定の値をとるため無視することができる．よって最終的に以下の式になる．

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \frac{1}{2}\int \widehat{w}(x)^2 P_s(\boldsymbol{x})dx - \int \widehat{w}(x)P_t(\boldsymbol{x})dx \\ &= \frac{1}{2}\sum_{l,l'=1}^{b} \boldsymbol{\alpha}_l \boldsymbol{\alpha}_{l'}\bigl(\int \varphi_l(\boldsymbol{x})\varphi_{l'}(\boldsymbol{x})P_s(\boldsymbol{x})dx\bigr) - \sum_{l=1}^{b} \boldsymbol{\alpha}_l\bigl(\int \varphi_l(\boldsymbol{x})P_t(\boldsymbol{x})dx\bigr) \\ &= \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{H}\alpha - \boldsymbol{h}^{\mathrm{T}}\alpha \end{aligned} \tag{6}$$

$\boldsymbol{H}$ は，$b * b$ の行列,

$$\boldsymbol{H}_{l,l'} = \int \varphi_l(\boldsymbol{x})\varphi_{l'}(\boldsymbol{x})P_s(\boldsymbol{x})dx \tag{7}$$

$\boldsymbol{h}$ は，b 次元のベクトル,

$$\boldsymbol{h}_l = \int \varphi_l(\boldsymbol{x})P_t(\boldsymbol{x})dx \tag{8}$$

である．このとき，標本平均をとり J の期待値を推定する．

$$\widehat{J}(\boldsymbol{\alpha}) = \frac{1}{2n_s}\sum_{i=1}^{n_s}\widehat{w}(x_i^s)^2 - \sum_{j=1}^{n^t}\widehat{w}(x_j^t) \tag{9}$$

$$= \frac{1}{2}\sum_{l,l'=1}^{b}\boldsymbol{\alpha}_l\boldsymbol{\alpha}_{l'}\Big(\frac{1}{n_s}\sum_{i=1}^{n_s}\varphi_l(x_i^s)\varphi_{l'}(x_i^s)\Big) - \sum_{l=1}^{b}\boldsymbol{\alpha}_l\Big(\frac{1}{n_t}\sum_{j=1}^{n_t}\varphi_l(x_j^t)\Big) \tag{10}$$

$$= \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^{\mathrm{T}}\boldsymbol{\alpha} \tag{11}$$

$\widehat{\boldsymbol{H}}$ は $b*b$ の行列,

$$\widehat{\boldsymbol{H}}_{l,l'} = \frac{1}{n_s}\sum_{i=1}^{n_s}\varphi_l(x_i^s)\varphi_{l'}(x_i^s) \tag{12}$$

$\widehat{\boldsymbol{h}}$ は b 次元のベクトル,

$$\widehat{\boldsymbol{h}}_l = \frac{1}{n_t}\sum_{j=1}^{n_t}\varphi_l(x_j^t) \tag{13}$$

である．最後に制約条件を考えない最適化問題の形に整えると以下のような式になる．

$$\min_{\alpha\in R}\Big[\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^{\mathrm{T}}\boldsymbol{\alpha} + \frac{\boldsymbol{\lambda}}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}\Big] \tag{14}$$

この無制約下での凸２次計画問題を解くことで $\boldsymbol{\alpha}$ を得ることができる．

4　実験結果

本実験では，現代日本語書き言葉均衡コーパス（BCCWJ コーパス (Maekawa (2007))）の OC(Yahoo! 知恵袋)，PB(書籍)，PN(新聞) の 3 つの異なる領域を扱い実験を行った．それぞれの領域から共に頻度が 50 以上多義語 16 単語を WSD の対象単語とする．領域適応としては，OC → PB，PB → OC，PB → PN，PN → PB，OC → PN，PN → OC の 6 通りの実験を行う．

分類器は Classias (Okazaki (2009)) を用いて学習した．(OC → PB) の結果を表 1 に，(PB → OC) の結果を表 2 に，(PB → OC) の結果を表 3 に，(PN → PB) の結果を表 4 に，(OC → PN) の結果を表 5 に，(PN → OC) の結果を表 6 に示す.

また，各表の列 Normal は確率密度比の重みを”1”とした時の結果，列 NB は NaiveBayes によって求めた確率密度比を適用した結果，uLSIF は uLSIF によって求めた確率密度比を適用した結果を表すものとする．Normal，NB，uLSIF の 3 つの手法で各単語ごとに領域適応を行い，その識別精度の平均値で評価する．

結果より，uLSIF により推定した確率密度比を適用しても正解率の改善が見込めることはなかった．

表 1: 実験結果 (OC → PB)

単語	Normal	NB	uLSIF
聞く	0.6667	0.6098	0.7642
言う	0.8205	0.7971	0.5835
子供	0.2903	0.7742	0.2473
時間	0.9054	0.9054	0.9054
自分	0.9513	0.8506	0.974
入れる	0.6727	0.5636	0.6364
出る	0.5658	0.6053	0.5592
取る	0.2593	0.2716	0.2716
場合	0.8467	0.854	0.8467
入る	0.4746	0.3729	0.4237
前	0.7438	0.8063	0.6438
見る	0.8493	0.8456	0.8382
持つ	0.7712	0.6993	0.6993
やる	0.9355	0.9355	0.9355
いく	0.8872	0.8647	0.8647
書く	0.7903	0.7903	0.9032
平均	0.7144	0.7216	0.6935

表 2: 実験結果 (PB → OC)

単語	Normal	NB	uLSIF
聞く	0.6774	0.5484	0.6613
言う	0.8363	0.8498	0.5315
子供	0.1558	0.1299	0.1299
時間	0.8302	0.8302	0.8302
自分	0.875	0.875	0.875
入れる	0.726	0.7534	0.7397
出る	0.6641	0.6183	0.687
取る	0.2787	0.3607	0.0984
場合	0.9286	0.9048	0.9762
入る	0.6029	0.5588	0.6324
前	0.8942	0.8846	0.875
見る	0.5611	0.5573	0.5611
持つ	0.8033	0.8197	0.5902
やる	0.9402	0.9402	0.9402
いく	0.6895	0.6119	0.6849
書く	0.7374	0.7374	0.7374
平均	0.7	0.6863	0.6594

表 3: 実験結果 (PB → PN)

単語	Normal	NB	uLSIF
聞く	0.7115	0.6923	0.7115
言う	0.9366	0.9284	0.7218
子供	0.5517	0.5517	0.5172
時間	0.661	0.661	0.661
自分	0.9859	0.9859	0.9859
入れる	0.7188	0.7813	0.7188
出る	0.8427	0.7753	0.7191
取る	0.4186	0.3721	0.2791
場合	0.8493	0.8493	0.8493
入る	0.5692	0.5538	0.2154
前	0.875	0.8558	0.7308
見る	0.7011	0.7586	0.6897
持つ	0.9153	0.8814	0.9153
やる	0.963	0.963	0.963
いく	0.8519	0.8148	0.8519
書く	0.7778	0.7037	0.7778
平均	0.7706	0.758	0.7067

表 4: 実験結果 (PN → PB)

単語	Normal	NB	uLSIF
聞く	0.5528	0.2602	0.7642
言う	0.8402	0.7828	0.5835
子供	0.5161	0.2473	0.2473
時間	0.8649	0.9054	0.9054
自分	0.974	0.974	0.974
入れる	0.6545	0.4	0.6364
出る	0.6447	0.5592	0.5592
取る	0.3827	0.2716	0.4074
場合	0.8394	0.1533	0.8467
入る	0.5897	0.3761	0.3761
前	0.8938	0.8562	0.6438
見る	0.8519	0.8444	0.8444
持つ	0.6993	0.6993	0.6993
やる	0.9416	0.9416	0.9416
いく	0.8647	0.8647	0.8647
書く	0.8387	0.9032	0.9032
平均	0.7468	0.6275	0.6998

表 5: 実験結果 (OC → PN)

単語	Normal	NB	uLSIF
聞く	0.4231	0.4615	0.7115
言う	0.4711	0.4545	0.7218
子供	0.4483	0.5172	0.4828
時間	0.678	0.661	0.661
自分	0.9718	0.9296	0.9859
入れる	0.7188	0.7188	0.7188
出る	0.7079	0.7416	0.7191
取る	0.2558	0.2558	0.2326
場合	0.8493	0.8493	0.8493
入る	0.3231	0.4154	0.2154
前	0.7619	0.7238	0.7238
見る	0.6897	0.7241	0.6897
持つ	0.8983	0.9153	0.9153
やる	1	1	1
いく	0.8519	0.8519	0.8519
書く	0.8148	0.8148	0.7778
平均	0.679	0.6897	0.7035

表 6: 実験結果 (PN → OC)

単語	Normal	NB	uLSIF
聞く	0.5242	0.3629	0.6613
言う	0.8318	0.7553	0.5315
子供	0.5325	0.8701	0.8701
時間	0.8302	0.8302	0.8302
自分	0.875	0.875	0.875
入れる	0.7534	0.2877	0.7397
出る	0.6947	0.5878	0.687
取る	0.2623	0.0492	0.0984
場合	0.9603	0.9762	0.9762
入る	0.4328	0.3433	0.2985
前	0.8762	0.7905	0.8667
見る	0.5692	0.5654	0.5654
持つ	0.5902	0.5574	0.5902
やる	0.9565	0.9565	0.9565
いく	0.6849	0.6849	0.6849
書く	0.6869	0.7374	0.7374
平均	0.6913	0.6394	0.6856

また，各領域適応の平均正解率を表 7 に，各手法の全領域適応の平均正解率の平均を評価値として表 8 に示す.

表 7: 各領域適応の平均値

領域適応	Normal	NB	uLSIF
OC → PB	0.7144	0.7216	0.6935
PB → OC	0.7	0.6863	0.6594
PB → PN	0.7706	0.758	0.7067
PN → PB	0.7468	0.6275	0.6998
OC → PN	0.679	0.6897	0.7035
PN → OC	0.6913	0.6394	0.6856

表 8: 各手法の評価値

手法	評価値
Normal	0.717
NB	0.6871
uLSIF	0.6916

5　考察

5.1　NB と uLSIF の比較

本実験で算出した NB と uLSIF の値を比較するために，各領域適応の各対象単語に関して両者の手法で算出した確率密度比が小さいデータ上位 10% を取り出した．NB により取り出したデータ群を H とし，uLSIF により取り出したデータ群を K として，$|H \cap K|/|H \cup K|$ の値を調べた．この値が 1 に近ければ NB と uLSIF の値は概ね同じような傾向があると言える．結果を表 9 に示す.

表 5.1 をみると明らかに NB と uLSIF は異なる値を付与している．どちらの値がより真の値に近いかは判断できないが，算出される確率密度比はかなり異なる.

表 9: NB と uLSIF の密度比が低い共通用例

単語	OC → PB	PB → OC	PB → PN	PN → PB	PN → OC	OC → PN
聞く	0.6	0.333	0.143	0.667	0.667	0.263
言う	0.158	0.121	0.099	0.241	0.108	0.128
子供	0.273	0.286	0	1	1	0.077
時間	0.429	0.167	0.4	0.2	0.2	0.429
自分	0.333	0.132	0.034	0.556	0.4	0.091
入れる	0.4	0.111	0.111	1	0.5	0.077
出る	0.3	0.111	0.304	0.385	0.2	0.529
取る	0.2	0.333	0	0.333	0.333	0.091
場合	0.412	0.13	0.13	0.556	0.556	0.2
入る	0.2	0.1	0.294	0.333	0.333	0.5
前	0.333	0.333	0.185	0.333	0.333	0.25
見る	0.083	0.256	0.059	0.455	0.455	0
持つ	0.333	0.2	0.071	0.333	0.5	0.333
やる	0.375	0.154	0	0.333	0.333	0
いく	0.257	0.182	0	0	0.333	0
書く	0.053	0.333	0.091	0	1	0.053
平均	0.429	0.258	0.071	0.333	0.5	0.132

5.2 確率密度比の p 乗の利用

共変量シフト下での学習ではパラメータ推定のバラツキが大きいことが指摘されている．このパラメータのバラツキによって識別精度が安定しないと考えられる．そこで，確率密度比 $w(\boldsymbol{x})$ を $0 < p < 1$ 乗した $w(\boldsymbol{x})^p$ を重みとして適用し，バラツキを抑えることが提案されている (杉山将 (2006))．

ここで以上のバラツキを抑える手法を用いて再度実験を行ってみる．前章の実験において NB と uLSIF で，確率密度比を $w(\boldsymbol{x})^{0.5}$ として実験を行う．その結果を表 10 と表 11 に示す．

表 10: 実験結果

領域適応	Normal	NB	uLSIF	NB ($w(\boldsymbol{x})^{0.5}$)	uLSIF ($w(\boldsymbol{x})^{0.5}$)
OC → PB	0.7144	0.7216	0.6935	0.712	0.7038
PB → OC	0.7	0.6863	0.6594	0.6849	0.659
PB → PN	0.7706	0.758	0.7067	0.7595	0.7203
PN → PB	0.7468	0.6275	0.6998	0.6484	0.7049
OC → PN	0.679	0.6897	0.7035	0.6929	0.6829
PN → OC	0.6913	0.6394	0.6856	0.6767	0.7075

NB も uLSIF も平方根を取ることで全体の評価値は向上しているがわずかに uLSIF の方がその値は高く，確率密度比を直接推定した方がより有効であることがここからもわかる．ただしどちらにしても Normal よりは評価値は低かった．

表 11: 各手法の評価値

手法	評価値
Normal	0.717
NB	0.6871
uLSIF	0.6916
NB ($w(\boldsymbol{x})^{0.5}$)	0.6957
uLSIF ($w(\boldsymbol{x})^{0.5}$)	0.6964

6　おわりに

本論文では，WSD の領域適応の問題が共変量シフトの問題であることを示し，共変量シフトの解法である確率密度比を重みとしたパラメータ学習により問題の解決をを図った．共変量シフト下での学習では，訓練データ $(\boldsymbol{x}, y)$ に対して，確率密度比 $w(\boldsymbol{x}) = \frac{P_t(\boldsymbol{x})}{P_s(\boldsymbol{x})}$ の重みをつけて学習する．ここでは $P_t(\boldsymbol{x})$ と $P_s(\boldsymbol{x})$ をそれぞれ求め，その比を取る手法と，直接 $w(\boldsymbol{x})$ を推定する手法である uLSIF を試した．どちらの手法も用いても識別の精度は向上しなかった．またそれぞれの手法の値を比較すると，かなり異なる値であることも確認できた．どちらがより有効な推定方法であるかは，今後調査する必要がある．

文献

Daumé III, Hal (2007) “Frustratingly Easy Domain Adaptation,” in *ACL-2007*, pp. 256–263.

Jing Jiang and Chengxiang Zhai (2007) “Instance weighting for domain adaptation in NLP,” in *ACL-2007*, pp. 264–271.

Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama (2009) “A Least-squares Approach to Direct Importance Estimation,” in *Journal of Machine Learning Research 10*, pp. 1391–1445.

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

Naoaki Okazaki (2009) “Classias: a collection of machine-learning algorithms for classification.”

Hidetoshi Shimodaira (2000) “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *Journal of statistical planning and inference*, Vol. 90, No. 2, pp. 227–244.

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

Hiroya Takamura Yosuke Saiki and Manabu Okumura (2008) “Domain Adaptation in Sentiment Classification by Instance Weighting,” *IPSJ SIG Technical Report. SIG-NL Report*, Vol. 2008, No. 33, pp. 61–67.

高村大也 (2010) 言語処理のための機械学習入門，コロナ社.

新納浩幸、佐々木稔 (2014) 「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」，自然言語処理，第 21 巻，第 1 号，(to appear).

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，第 13 巻，第 3 号，pp.111–118.